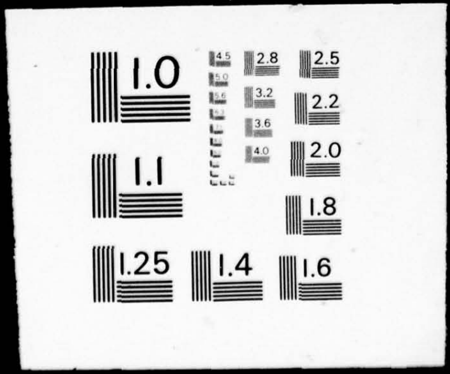
AD-A031 017   AUBURN UNIV ALA SCHOOL OF ENGINEERING   F/G 5/2
1975 USAF/ASEE SUMMER FACULTY RESEARCH PROGRAM.(U)
SEP 75 J F O'BRIEN   F44620-75-C-0031
UNCLASSIFIED   AFOSR-TR-76-1113   NL

1 OF 7
ADA031017

1 OF 7

ADA031017



2

FG.

AFOSR - TR - 76 - 1113

1975 USAF-ASEE

SUMMER FACULTY RESEARCH PROGRAM

Research Reports

Conducted by:
SCHOOL OF ENGINEERING
AUBURN UNIVERSITY
Auburn, Alabama
September, 1975

Approved for public release;
distribution unlimited.

6

1975 USAF/ASEE Summer Faculty RESEARCH PROGRAM.

9 Final rept,

Conducted by
Auburn University
with Assistance from
Ohio State University
at
Wright Patterson and Eglin Air Force Bases
under
USAF Contract Grant F 44620-75-C-0031

15

16 AF-9768 17 976802

PARTICIPANT'S RESEARCH REPORTS

18 AFOSR 19 TR-76-1113

Submitted to
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, D. C.

10

by
J. Fred O'Brien, Jr., University Project Director
Associate Director, Engineering Extension Service
Auburn University

11 September 1975

12 579p.

1473

390276

4B

PREFACE

The 1975 USAF-ASEE Summer Faculty Research Program consisted of twenty-two members of engineering and science faculties from twenty colleges and universities throughout the nation, engaged in scientific research of mutual interest and benefit to the Associate (and his university) and the USAF. These professors were assigned to various USAF research laboratories at Wright-Patterson AFB and Eglin AFB for a ten-week period of concentrated research in their selected field in collaboration with an assigned USAF colleague.

The basic program objectives were:

(1) To provide scientific and technological benefits to the USAF while enhancing the research interests and capabilities of engineering educators.

(2) To stimulate continuing relations among participating faculty members and their professional peers at the AFSC laboratories.

(3) To form the basis for continuing research of interest to the Air Force at the participant's institution.

(4) To sponsor research in areas of mutual interest to the USAF, the faculty member, and his institution.

The program was conducted under contract with the Air Force by the School of Engineering, Auburn University with assistance from the School of Engineering, Ohio State University. The American Society for Engineering Education is co-sponsor of the program

This document is a compilation of each associate's report assembled by J. Fred O'Brien, Jr., Project Director, who has exercised certain administrative prerogatives to produce this report.

List of 1975 Participants

Dr. John A. Alic
Ph.D., Assistant Professor, 33
Mechanical Engineering, Wichita State University
USAF Assignment: Structural Integrity
Structures
Flight Dynamics
Research Topic : Fatigue Crack Propagation in Laminated and Monolithic Aluminum Alloy Panels
USAF Colleague : J. P. Gallagher

Dr. Louis I. Boehman
Ph.D., Associate Professor, 37
Mechanical Engineering, University of Dayton
USAF Assignment: Aero Propulsion Laboratory
Fuels and Lubrication
Fuels

Research Topic : Design of a Subsonic Combustion Tunnel and Experimental test program
USAF Colleague : William F. Blazowiski Ph.d.

Dr. J. Kent Bryan
Ph.D., Assistant Professor, 32
Electrical and Computer Engineering, Clemson University
USAF Assignment: Nondestructive Evaluation Br.
Metals and Ceramics Division
Air Force Materials Laboratory
Research Topic : Pattern Recognition Techniques Applied to Flat-Bottom Holes
USAF Colleague : M. J. Buckley

Dr. James F. Delansky
Ph.D., Associate Professor, 41
Electrical Engineering, Penn State University
USAF Assignment: Systems Analysis and Simulation
Digited Guided Weapons
Armament Laboratory
Research Topic : Digital Autopilot Design

USAF Colleague : Major K. A. (Al) Gale

Dr. M. Paul Hagelberg
Ph.D., Professor, 42
Physics, Wittenberg University
USAF Assignment: Nondestructive Evaluation
Metals and Ceramics
Materials
Research Topic : Ultrasonic Techniques for Nondestructive Evaluation (NDE)
USAF Colleague : Dennis Corbly

Dr. Bruce P. Johnson
Ph.D., Associate Professor, 37
Electrical Engineering, University of Nevada
USAF Assignment: Microwave Technology
Electronic Technology
Avionics
Research Topic : Ohmic Contacts for Transferred Electron Devices
USAF Colleague : G. L. McCoy/C. I. Huang

Dr. Joseph T. Maloy
Ph.D., Assistant Professor, 36
Chemistry, West Virginia University
USAF Assignment: Energy Conversion
Aerospace Power
Aero Propulsion
Research Topic : The Effect of Cobalt Hydroxide Coprecipitation in Nickel Hydroxide Electrodes
USAF Colleague : David F. Pickett

Dr. D. Maples
Ph.D., Associate Professor, 38
Mechanical Engineering-Thermal Science, Louisiana State University
USAF Assignment: Aircraft Compatibility
Munitions
Armament
Research Topic : Transient Thermal Analysis of External Stores
USAF Colleague : J. C. Key, Jr.

Dr. Richard A. Miller
Ph.D., Assistant Professor, 31
Industrial and Systems Engineering, Ohio State University
USAF Assignment: Systems Evaluation and Systems Technology
Human Engineering and Environmental Medicine
Aerospace Medical Research Lab
Research Topic : Continuous Performance Measurement of Manually Controlled Systems
USAF Colleague : Jerry P. Chubb/Carroll N. Day

Dr. Don H. Morris
Ph.D., Assistant Professor, 36
Mechanical Engineering, Mississippi State University
USAF Assignment: Mechanics and Surface Interactions
Nonmetallic Materials
Materials
Research Topic : Fracture of Graphite/Epoxy Composites
USAF Colleague : Dr. H. T. Hahn

Dr. Philip S. Noe
Ph.D., Assistant Professor, 44
Electrical Engineering, Texas A & M University
USAF Assignment: Analysis
Reconnaissance & Weapon Delivery
Avionics
Research Topic : A Navigation Algorthm for the Low-Cost GPS Receiver
USAF Colleague : K. A. Myers/ D. Botha

Dr. Charles E. Nuckolls

Ph.D., Assistant Professor, 39
Mechanical Engineering, Florida Technological University
USAF Assignment: Recovery and Crew Station
Vehicle Equipment
Flight Dynamics
Research Topic : RPV Ground Impact Attenuator Simulation
USAF Colleague : R. Harley Walker

Dr. Jerry W. Rogers

Ph.D., Associate Professor, 44
Electrical Engineering, Mississippi State University
USAF Assignment: Analysis and Evaluation
Reconnaissance and Weapon Delivery
Avionics
Research Topic : Electro-Optical Tracker Analysis
USAF Colleague : Capt. Gary Reid

Dr. Philip C. Rymers

Ph.D., Professor, 46
Mechanical Engineering, University of Nevada
USAF Assignment: Aerospace Dynamics
Vehicle Dynamics
Air Force Flight Dynamics
Research Topic : A Method for Investigating the Angular Vibration
Response of a Structure
USAF Colleague : R. N. Bingman

Dr. Charles R. Slivinsky

Ph.D., Associate Professor, 34
Electrical Engineering, University of Missouri-Columbia
USAF Assignment: Control Systems Development
Flight Control
Flight Dynamics
Research Topic : Analysis of Inherent Errors in Asynchronous Redundant
Digital Flight Controls
USAF Colleague : Capt. Vincent J. Darcy

Dr. Thomas G. Stoebe

Ph.D., Associate Professor, 36
Mining, Metallurgical & Ceramic Engineering, University of Washington
USAF Assignment: Laser & Optical Meterials
Electromagnetic Meterials
Materials
Research Topic : Optical Properties of Europium-Doped Pottasium Chloride
Laser Window Materials
USAF Colleague : G. Edward Kuhl

Dr. Donald C. Stouffer

Ph.D., Associate Professor, 37
Engineering Analysis, University of Cincinnati
USAF Assignment: Metals Behavior
Metals and Ceramics
Materials Laboratory
Research Topic : An Analysis of Varying Material Properties: A Measure of
Damage
USAF Colleague : T. Nicholas

Dr. Thomas A. Stuart
Ph.D., Assistant Professor, 34
Electrical Engineering, Clarkson College
USAF Assignment: Power Distribution
Aerospace Power
Aero Propulsion
Research Topic : Overload Protection and Filterinq Requirements for
Phase Control Voltage Regulators
USAF Colleague : P. C. Herren/P. E. Stover

Dr. Shu-Yi S. Wang
Ph.D., Associate Professor, 39
Mechanical Engineering, University of Mississippi
USAF Assignment: Components
Turbine Engine
AF Aero Propulsion
Research Topic : A Study on Numerical Methods for Computing Transonic
Flows in Turbomachines
USAF Colleague : Kervyn D. Mach

Dr. Michael E. Warren
Ph.D., Assistant Professor, 28
Electrical Engineering, University of Florida
USAF Assignment: Mines
Minitions
Armament
Research Topic : Analysis of Missile Control Systems
USAF Colleague : Major L. D. Berry

Dr. Richard J. Wolf
Ph.D., Assistant Professor, 27
Industrial Engineering, New Mexico State University
USAF Assignment: Active ECM
Electronic Warfare
Avionics
Research Topic : A Real Time Terminal Guidance Simulation Facility
USAF Colleague : E. F. Mayleben

Dr. J. N. Youngblood
Ph.D., Associate Professor, 37
Aerospace and Mechanical Engineering, University of Alabama
USAF Assignment: Systems Analysis and Simulation
Digited Guided Weapons

Research Topic : Digital Autopilot Design
USAF Colleague: Major K. A. (Al) Hale

contents:

RESEARCH REPORTS

p. viii

RESEARCH REPORTS (continued)

Report No.	Title	Faculty Fellow
15	Fatigue Crack Propagation in Laminated and Monolithic Aluminum Alloy Panels	Dr. J. A. Alic
16	RPV Ground Impact Attenuator Simulation	Dr. Charles E. Nuckolls
17	A Method for Investigating the Angular Vibration Response of a Structure	Dr. Philip C. Rymers
18	Analysis of Inherent Errors in Asynchronous Redundant Digital Flight Controls	Dr. Charles Slivinsky
19	Pattern Recognition Techniques Applied to Flat-Bottom Holes	Dr. J. Kent Bryan
20	Ultrasonic Techniques for Nondestructive Evaluation (NDE)	Dr. M. Paul Hagelberg
21	Fracture of Graphite/Epoxy Composites	Dr. Don H. Morris
22	Optical Properties of Europium-Doped Pottasium Chloride Laser Window Materials	Dr. Thomas G. Stoebe

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO

&

EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

DESIGN OF A SUBSONIC COMBUSTION TUNNEL AND EXPERIMENTAL TEST PROGRAM. A NON-EQUILIBRIUM SOOTING LIMIT CRITERIA. EQUILIBRIUM FLAME TEMPERATURE COMPUTER PROGRAM

Prepared by:	Louis I. Boehman PhD.
Academic Rank:	Associate Professor
Department and University:	Department of Mechanical Engineering University of Dayton
Assignment:	
(Laboratory)	Aero Propulsion Laboratory
(Division)	Fuels and Lubrication
(Branch)	Fuels
USAF Research Colleague:	William S. Blazowski PhD.
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

DESIGN OF A SUBSONIC COMBUSTION TUNNEL AND EXPERIMENTAL TEST PROGRAMS. A NON-EQUILIBRIUM SOOTING LIMIT CRITERIA. AN EQUILIBRIUM FLAME TEMPERATURE COMPUTER PROGRAM

by

Louis I. Boehman

ABSTRACT

A continuous flow subsonic combustion tunnel to be used as a basic research tool was designed through the conceptual state up to the point of final engineering design. The tunnel incorporates capabilities for advanced optical combustion diagnostics (laser-Raman spectroscopy and laser-Doppler velocimetry). Plans for combustion experiments to be carried out in this new facility were developed. The experiments planned address topics of immediate and long range importance for combustion in air-breathing propulsion engines.

A second topic addressed in this report concerns the problem of predicting the conditions under which soot will form in pre-mixed fuel/air systems during combustion. A simple model for soot formation based on a common approach to balancing combustion reactions for fuel rich conditions was developed. This non-equilibrium model provides the maximum air-fuel ratio below which soot will form in pre-vaporized, premixed, short residence time combustors.

The third and final topic addressed in this report is the problem of calculating equilibrium flame temperatures. An efficient, easy to use, and reliable computer program was developed which calculates the equilibrium flame temperature and the equilibrium composition of the products of combustion obtained when a general C-H-O-N fuel is burned with air at constant pressure or constant volume. This computer program represents a modified version of an existing one developed previously by the author. The main modification was to add a subroutine which calculates relatively fool-proof initial guesses for the solution to the system of equations. The only input data required to operate the computer program are the fuel composition, the heating value of the fuel, the specific heat of the fuel, the initial temperature and pressure, specification of constant volume or constant pressure combination, and the air/fuel ratio.

SECTION 1

INTRODUCTION

Recent developments in terms of the necessity for clean combustion coupled with impending fuel shortages have served as the impetus for increased attention being paid to the design of air breathing propulsion engine combustors. The need for clean but yet efficient combustion in a gas turbine engine has prompted a new era in combustor design in which an analytical approach has largely displaced the old "cut and try" process. Likewise, the declining availability and increasing price of turbine engine fuel has dictated that alternate fuels be considered for use in turbine engines. These developments have severely tested the existing data bases of chemical kinetics, fluid dynamics, and numerical analysis. Progress in the development of computer codes for predicting the behavior of turbulent, reacting flows which are suitable for preliminary design of combustors has been slow because the generation of the fundamental data required to develop turbulence and chemical kinetic models has not kept pace with the need for such data.

These recent developments have spurred the development of new concepts for carrying out combustion in gas turbine engines and these new concepts have in turn created new problems to be addressed by the combustion designer. The three research topics covered in this report are all directed toward providing tools for evaluating new concepts and solving some of these new problems.

SECTION II

DESIGN OF A SUBSONIC COMBUSTION TUNNEL AND EXPERIMENTAL TEST PROGRAM

Non-flow perturbing measurement techniques have become available in recent years for combustion diagnostics. Laser Raman Spectroscopy (LRS) has been developed to the point where major species concentrations can be measured with LRS and Laser-Doppler Velocimetry (LDV) systems are now available "off the shelf". Both techniques allow one to make essentially point measurements in a reacting flow without disturbing the flow. Coherent Anti-Stokes Raman Spectroscopy (CARS) will undoubtedly be developed in just a few years to the point where minor species can be measured as well.

These techniques have all been used during the past five years or so to study open flames where the problem of allowing the optics to "see" the flame is trivial. The use of these techniques for combustion diagnostics in confined flows is obviously a more difficult problem, particularly in a high pressure, high temperature rise combustor.

In order to capitalize on the availability of LRS, LDV, and potentially in the near future, CARS, the Fuels Branch of AFAPL was interested in constructing a research combustion facility which incorporates capabilities for using these new optical combustion diagnostic tools. The design of this facility was thus undertaken.

The following requirements for the combustion tunnel were identified.

(1) The facility should utilize the full capabilities of the present air supply and air preheater facilities available in Building 18 with growth potential to include new air supply capabilities of a planned new facility.

(2) A high degree of flexibility is desired in order to provide capability for a maximum number of different types of experiments.

(3) Major emphasis to be placed on providing optical windows in the test section so that the entire combustion zone can be optically viewed and measured.

(4) The test section should be of rectangular cross-section and should be able to be mounted in both horizontal and vertical configurations.

(5) Both heterogenous and homogeneous modes of combustion are desired. Fuel preheating, pre-vaporization capability is required as well as premixing of fuel and air.

(6) Subatmospheric pressure capability should be provided in order to investigate combustion problems in afterburners. In order to meet all of the above diverse requirements, a modular design concept was chosen.

The following specific combustion research problems were identified as being important to the air-breathing combustion community.

(1) Catalytic combustion in premixed hydrocarbon air and hydrogen-air systems principally to minimize particulate and gaseous emission.

(2) Autoignition and flashback in premixed systems including development of mixing concepts to minimize the problem.

(3) Afterburner ignition using catalytic ignitors.

(4) Combustion of synthetic fuels formed from coal, oil shale, and tar sands.

(5) Determining the oxidation kinetics of higher hydrocarbons under both fuel-rich and fuel-lean conditions.

(6) Staged combustion processes.

(7) Combustion experiments in which initial and boundary conditions are carefully controlled in order to provide a data base for check out of computer simulation of combustion experiments (combustion models).

The conceptual design of a facility that has the potential to meet the requirements outlined above and which provides the capability of solving the above combustion problems is shown in Figure 1. A double walled tunnel was conceived in order to provide windows for viewing the combustion process under high temperature and pressure conditions. Combustion takes place in the inner duct only so that the windows in the inner ducts will experience high temperatures but no substantial pressure loading. The outer shell and its windows take virtually all of the pressure but will not experience high temperatures. It is proposed that the windows in the inner ducts be made of sapphire while fused quartz will probably be adequate for the outer windows.

A method for conveniently setting up the tunnel and the associated optical instrumentation is shown in Figure 2. It is proposed that a vertical milling machine be used as the support and 3-D traversing mechanism for the optics and lasers. The carriage drivers can easily be motorized so that the traverses can be done remotely. This is an absolute requirement from a safety point of view.

Detailed engineering design remains to be accomplished. It is suggested that the windows be designed with the aid of computer programs used for thermal and pressure stress analysis of windows used in high power lasers (Reference 1).

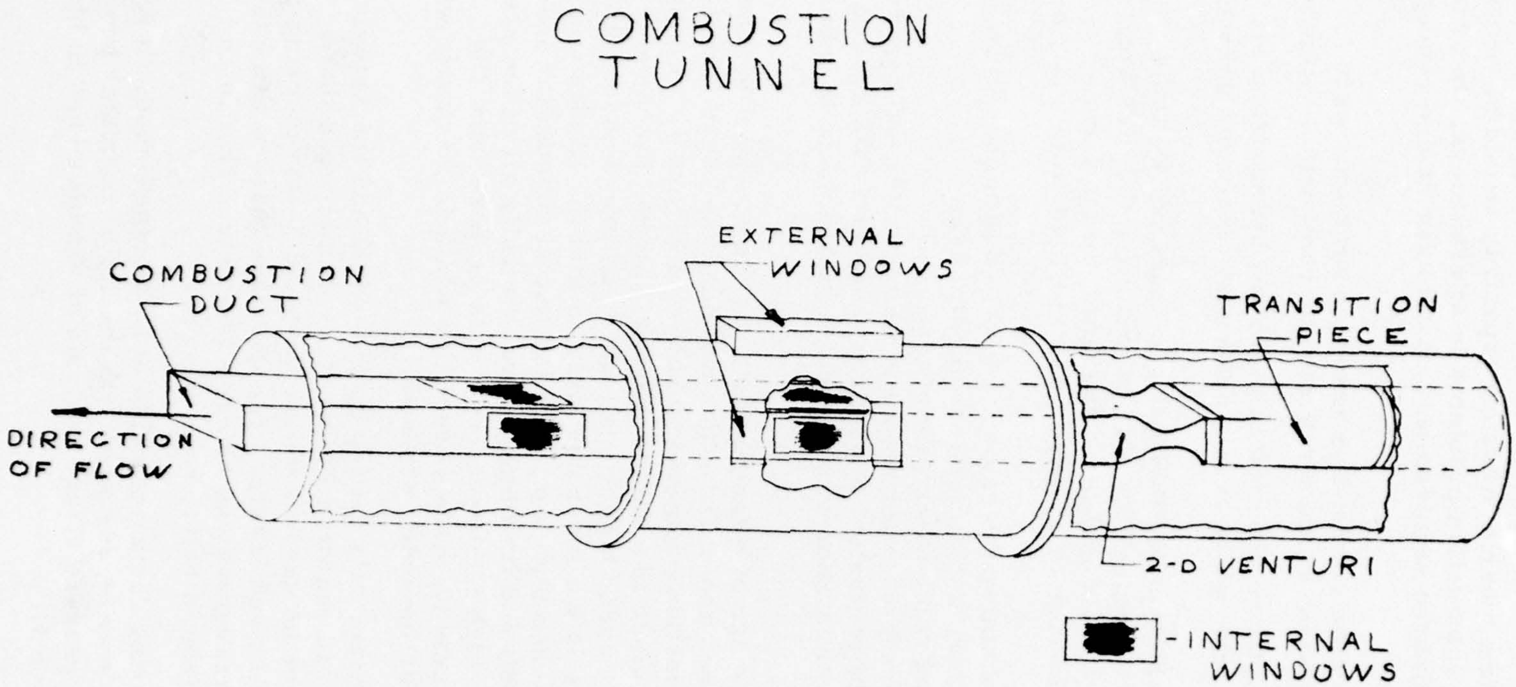


Figure 1. Sketch of Combustion Tunnel Showing Double Wall Construction

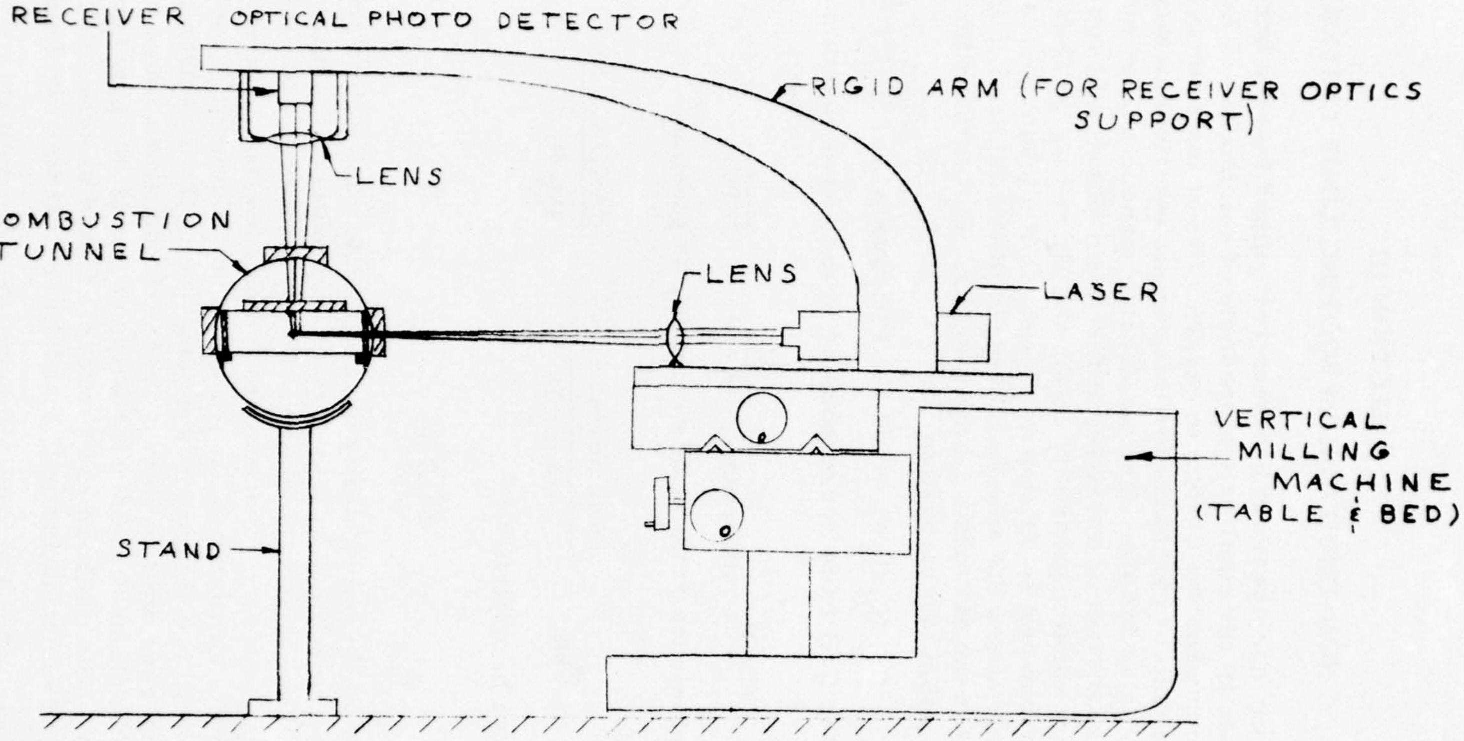

RECEIVER
OPTICAL PHOTO DETECTOR
RIGID ARM (FOR RECEIVER OPTICS SUPPORT)
LENS
COMBUSTION TUNNEL
LENS
LASER
VERTICAL MILLING MACHINE (TABLE & BED)
STAND

Figure 2. General Concept for Making Remote Measurements in Combustion Tunnel

SECTION III

A NON-EQUILIBRIUM SOOTING LIMIT CRITERIA

For many years it has been recognized that the onset of soot formation in the equilibrium products of combustion of hydrocarbon fuels will occur when the carbon to oxygen ratio of the reactants (C/O) is close to unity. The basis for this simple equilibrium sooting limit criteria is as follows: The equilibrium products of combustion of a hydrocarbon fuel at an oxidizer/fuel ratio where soot formation is imminent consist primarily of CO and H_2 and N_2. Thus, if a general hydrocarbon fuel is written as $C_M H_N Q_p N_Q$, soot formation will be imminent where the amount of oxygen in the oxidizer is just sufficient to convert the M carbon atoms into CO. The combustion reaction for this situation can be written as:

$$C_M H_N O_p N_Q + 1/2\ (M-P)O_2 \rightleftharpoons MCO + \frac{N}{2} H_2 + \frac{Q}{2} N_2$$

Thus, the C/O ratio corresponding to the above reaction is simply equal to

$$(C/O)_{equilibrium} = \frac{M}{M - P} = \frac{1}{1 - P/M} \tag{1}$$

The equivalence ratio at this condition is given by

$$\phi_{eq} = \frac{\text{stoichiometric } O_2}{\text{actual } O_2} = \frac{M + N/4 - P/2}{(M-P)\ /\ 2} \tag{2}$$

which can be written as

$$\phi_{eq} = 1 + \frac{M + N/2}{M - P} \tag{3}$$

A check on the applicability of this simple equilibrium sooting limit criteria is provided by comparison to results of detailed equilibrium thermochemistry calculations. In Figure 3 are shown the results of burning olefin or napthene hydrocarbons ($C_M H_{2M}$) in air. In Figure 4 are shown similar results for a number of other hydrocarbon fuels. From these two figures it is evident the C/O = 1 sooting limit criteria is a very good approximation over a temperature range of about 2160°R to 3960°R (1200°K to 2200°K) particularly at low pressures (less than 10 atm). The methods and procedures used to determine the sooting limits presented in Figures 3 and 4 are described in Reference 2.

The C/O = 1 sooting limit criteria is applicable to combustors which are essentially well-stirred reactors with relatively long residence

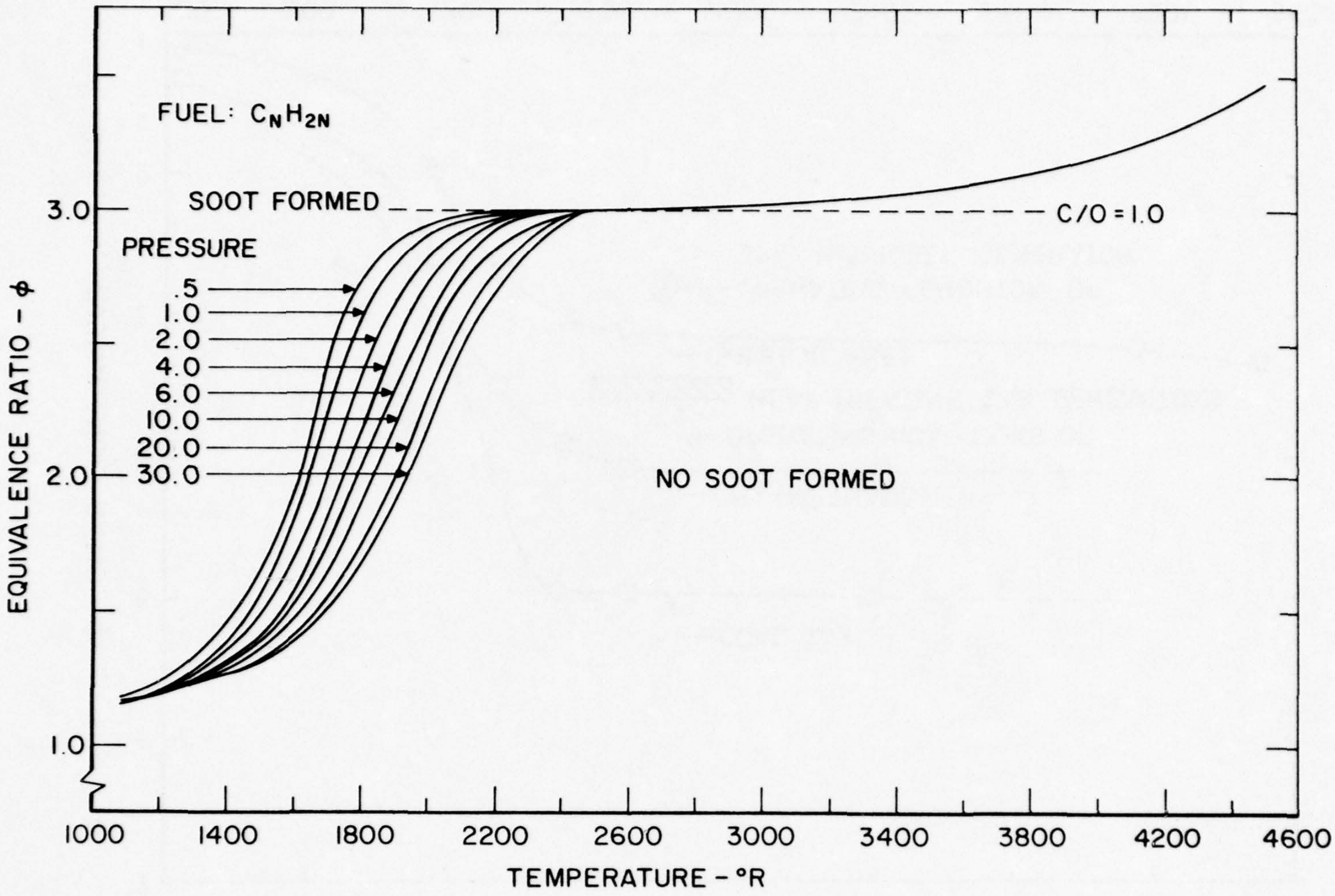


Figure 3. Equilibrium Sooting Conditions for $C_n H_{2n}$ Fuel

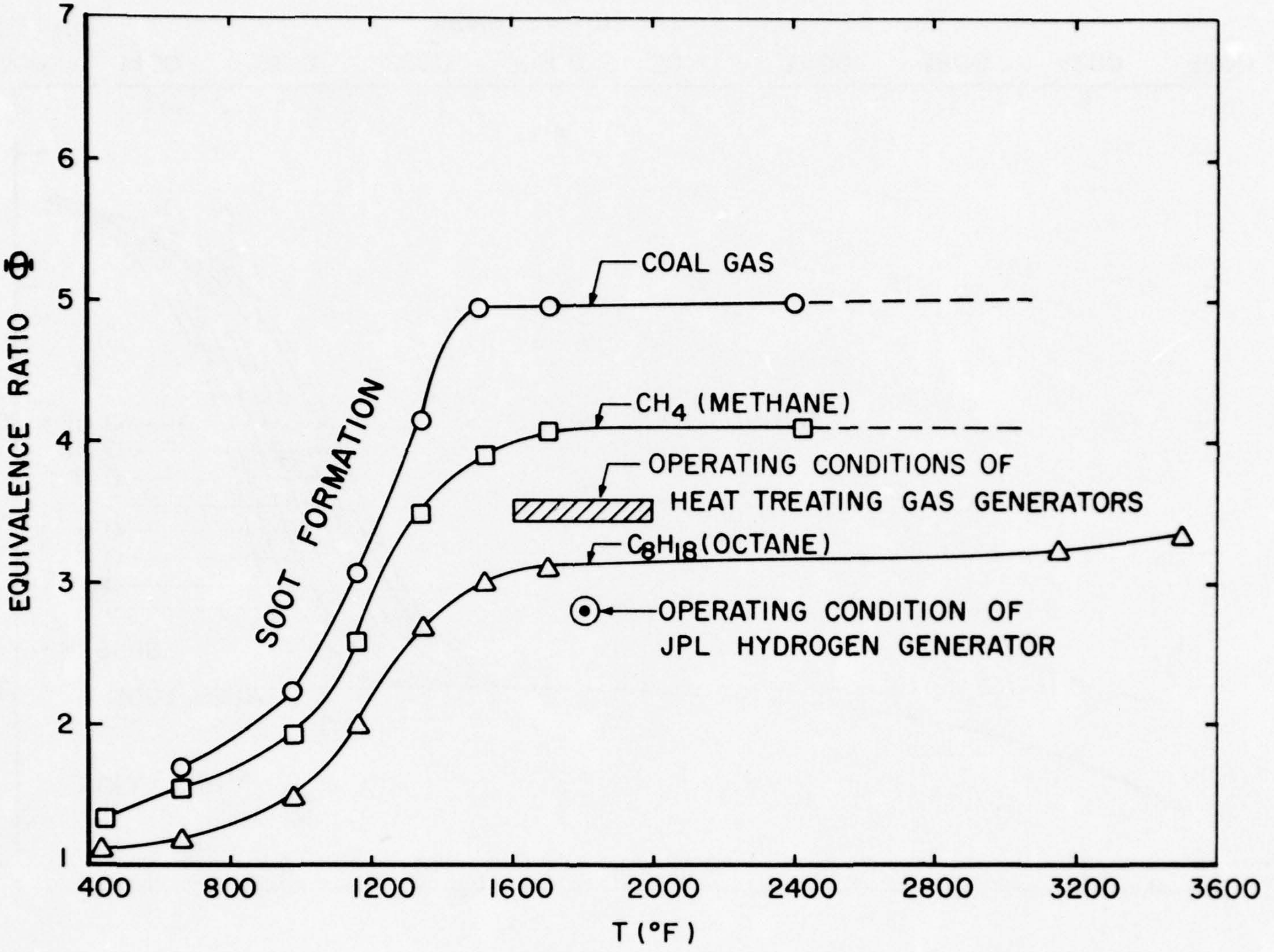


Figure 4. Equilibrium Sooting Conditions for Several Fuels

times. Many industrial combustors, particularly those in reducing atmosphere generators are designed on the basis of the C/O = 1 sooting limit criteria with good success. The residence time required to achieve equilibrium in a well stirred reactor decreases as the temperature increases and can be shorter than 3-4 milliseconds at temperatures in the neighborhood of 2000°K (Reference 3) and as long as 100 milliseconds at temperatures near 1000°K (Reference 4) for simple hydrocarbon fuels. In modern turbine propulsion engines 10 milliseconds is considered to be sufficient residence time for completion of the chemical reactions.

When the residence time is too short for equilibrium to occur, the composition of the products of combustion are very difficult to predict with strong dependence on the degree of premixing of the fuel and oxidizer and the macrostructure of the turbulence in the entire combustor. In spite of these difficulties, however, a lower limit to the (C/O) ratio below which soot will not occur can be determined very easily and appears to correlate very well with soot formation data for well mixed systems of some simple hydrocarbon fuels burned with oxygen or air. This model is based on the observation that in the flame zone, the kinetics of conversion of H_2 to H_2O is generally much faster than the conversion of carbon or carbon monoxide to CO_2. Under conditions when this assumption is valid, the following approach can be used. It will be assumed that soot formation will be imminent when just enough oxidizer is supplied to convert all of the carbon in the fuel to CO. Then the combustion reaction becomes:

$$C_M H_N O_P N_Q + 1/2(M + \frac{N}{2} - P)O_2 \longrightarrow MCO + \frac{N}{2} H_2O + \frac{Q}{2} N_2 \qquad (4)$$

This reaction is a special form of the combustion reaction obtained when combustion occurs with a deficiency of oxygen and a well known procedure for balancing the reaction is applied. In this procedure, it is assumed that hydrogen takes what oxygen it requires for complete combustion and CO and CO_2 share whatever oxygen remains. When this simple procedure is applied, it is found that when the oxygen supplied drops below a certain value, negative amounts of CO_2 are obtained. The above form of the reaction corresponds to the case where the amount of CO_2 is zero.

The (C/O) ratio for the above reaction (denoted by $(C/O)_{neq}$ is given by

$$(C/O)_{neq} = \frac{M}{M + \frac{N}{2} - P} = \frac{1}{1 + \frac{N}{2\,M} - \frac{P}{M}} \qquad (5)$$

and the corresponding equivalence ratio is given by

$$\phi_{neq} = 1 + \frac{1}{1+1/2\,\frac{N}{M} - \frac{P}{M}} \qquad (6)$$

which can also be written as

$$\phi_{neq} = 1+(C/O)_{neq} \qquad (7)$$

The applicability of the above equations for predicting the minimum (C/O) ratio above which soot may form is demonstrated by comparing the predictions of equation (5) with experimental data for burning acetylene, ethylene, and ethane as shown in Figures 5, 6, and 7 respectively. All of the data presented in these three figures are taken directly from figures in Reference 3 in which available data on soot formation in premixed flames for these three fuels was compared to data obtained by the authors of Reference 3 in a shock tube. The data given in Reference 3 for acetylene is shown in Figure 5. Equation 5 gives a value of $(C/O)_{neq}$ of .667 (ϕ_{neq} = 1.667) for C_2H_2 . It is observed that the flat flame data of Fenimore, et al. (FJM)[5] , of Homann and Wagner (HW)[6] and of Street and Thomas (ST)[7] all lie between the C/O=1 limit and the $(C/O)_{neq}$ limit of .667. The bar on the data of FJM includes the $(C/O)_c$ that they determined at all flame pressures in range 10 to 60 torr; the data of HW were taken at 20 torr; and the ST data were at 1 atm. The Radcliffe and Appelton (RA) shock tube data are all observed to lie above the C/O equilibrium limit. RA suggest that this high temperature "super equilibrium" phenomena is due to the presence of hydroxyl radical concentrations much greater than equilibrium which, following the suggestion of Millikan[8], inhibits the growth of soot particles. The RA suggestion is equilavent to stating that the carbon is present but exists either in a supersaturated nonequilibrium vapor state or bound to hydrogen in the form of polyacetylenes (Reference 9).

This "super equilibrium" phenomena is typical of what occurs in the region just following the flame front ie., just following the oxidation region. Since RA believe that their data is applicable to this region, it is clear that the RA data are not really comparable to data from flame or reactor experiments where the residence times are sufficient to allow soot formation in the burned gas zone (the zone downstream of the highly non-equilibrium oxidation zone).

In Figure 6 the data given in Reference 3 for ethylene is shown. Equation 5 gives a value of $(C/O)_{neq}$ of 0.5 (ϕ_{neq} = 1.5) for C_2H_4 . Here again it is observed that the flame data all lie between the $(C/O)_{neq}$ limit and the $(C/O)_{eq}$=1 limit. The flame measurements shown are due

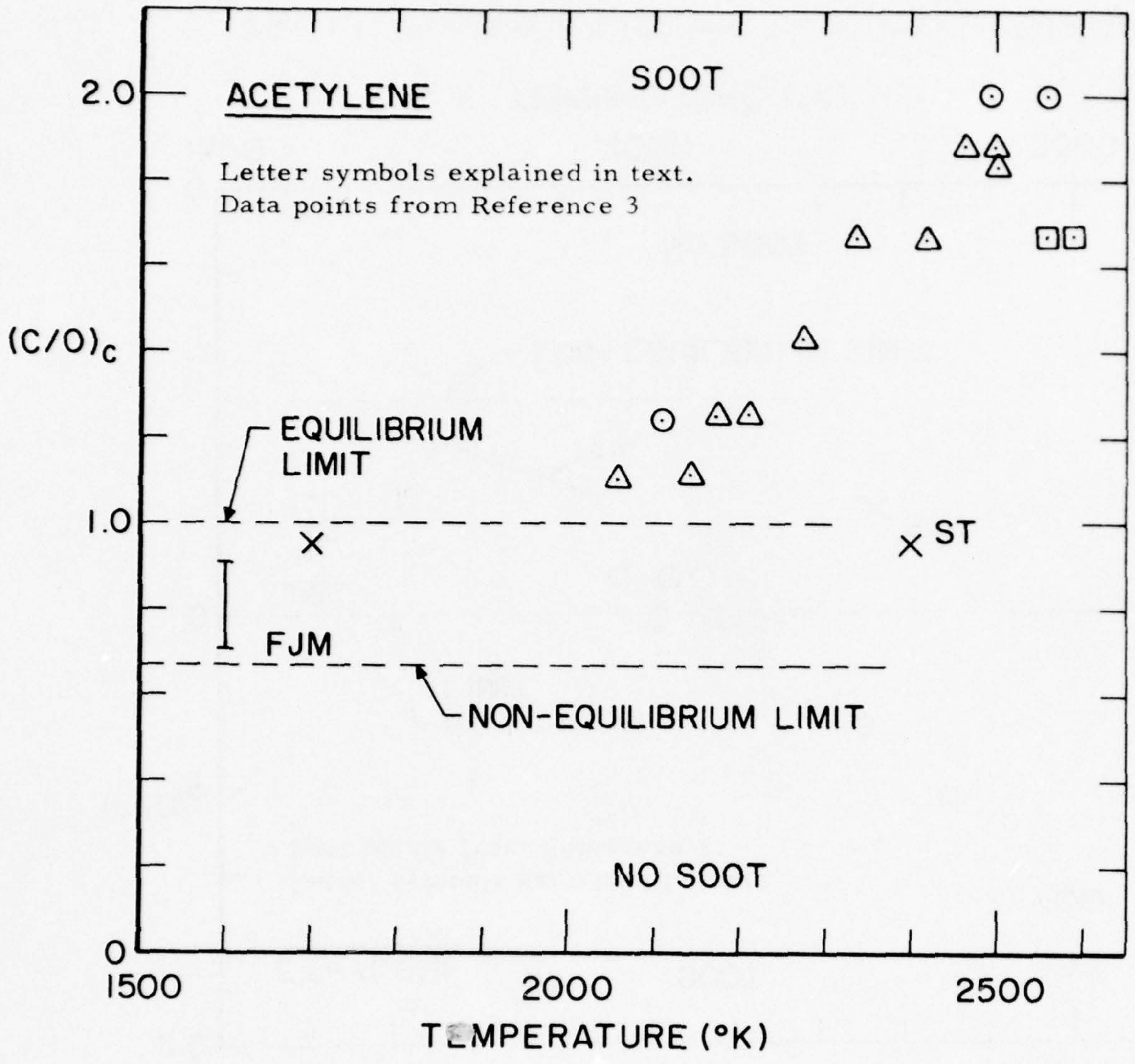


Figure 5. Variation of $(C/O)_c$ with Temperature for Acetylene.

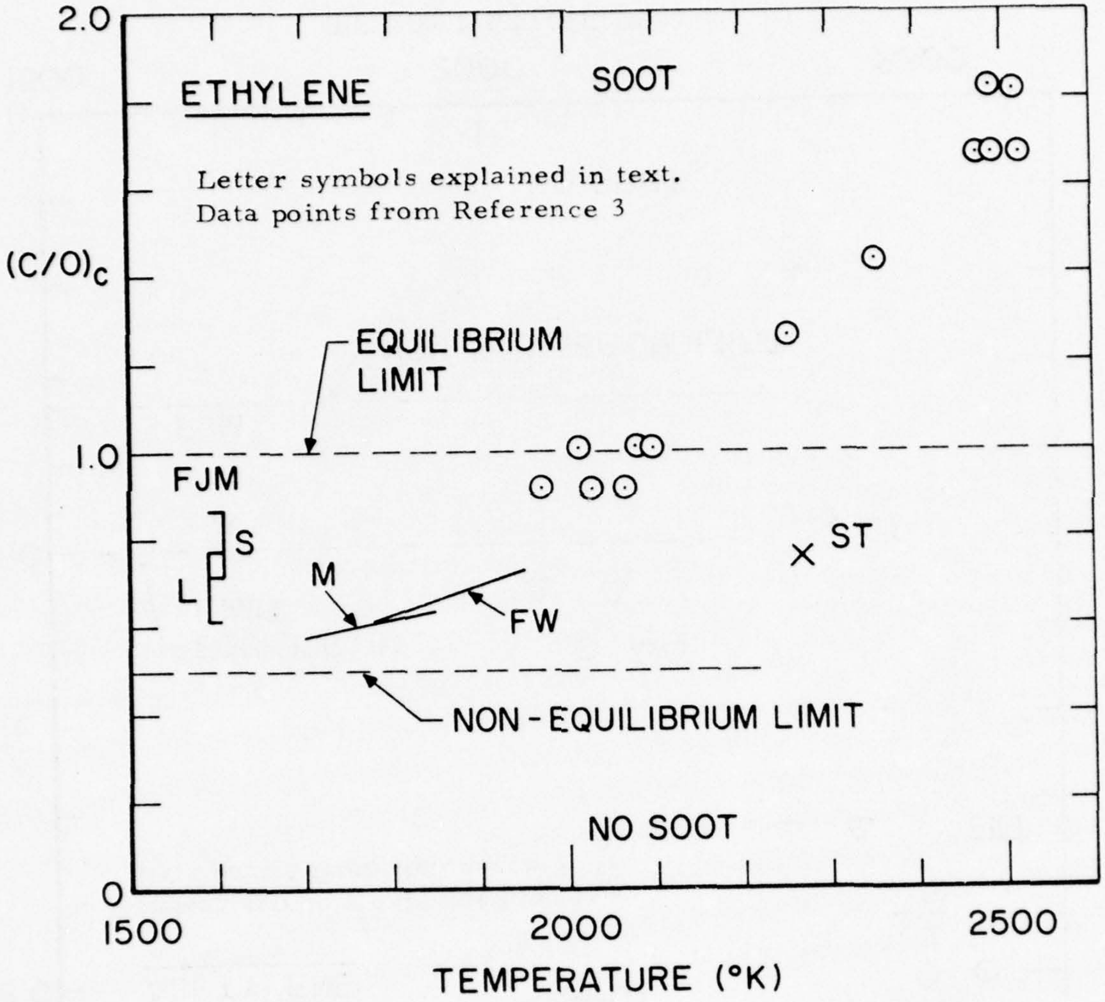


Figure 6. Variation of $(C/O)_c$ with Temperature for Ethylene.

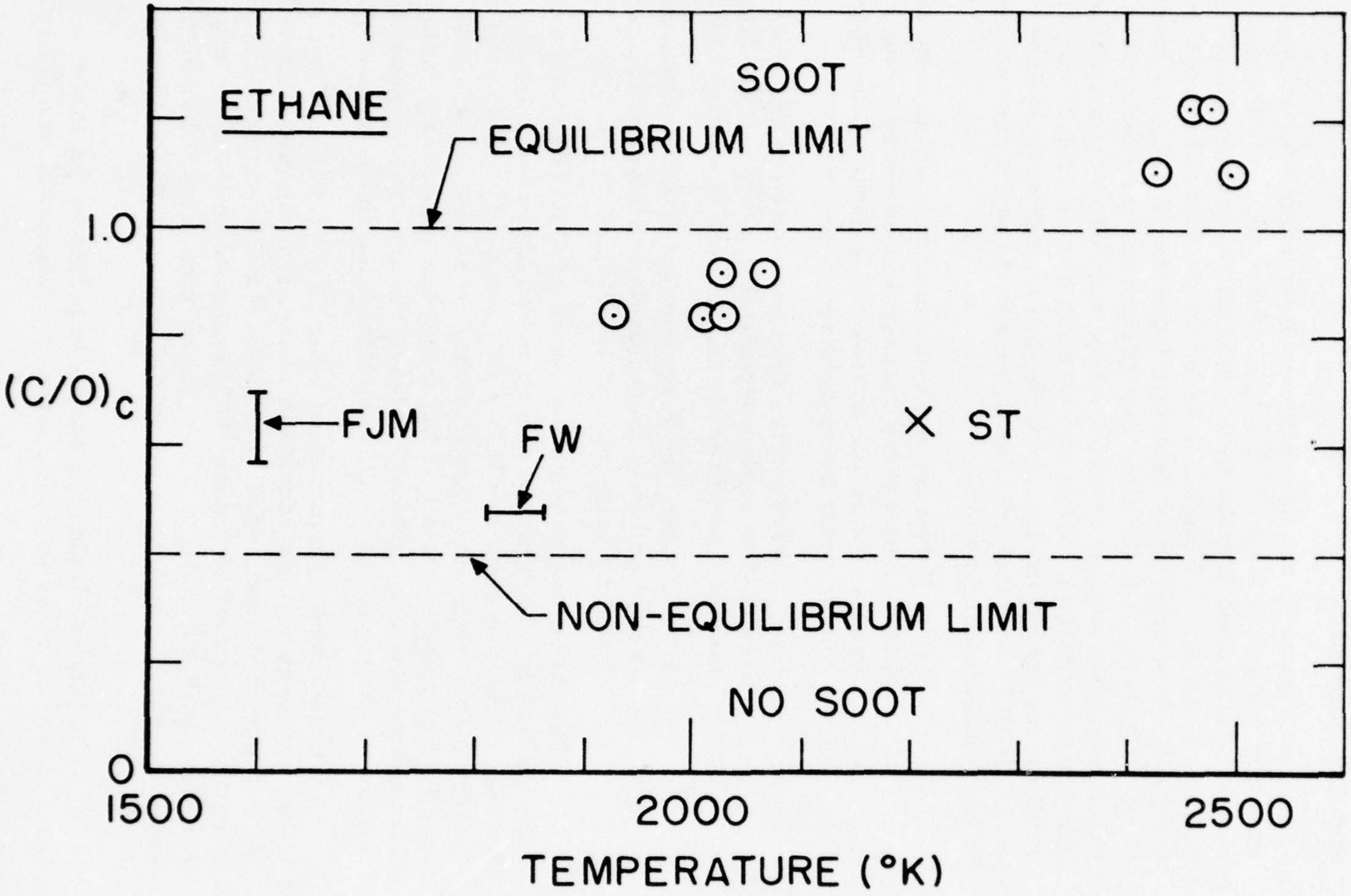


Figure 7. Variation of $(C/O)_c$ with Temperature for Ethane.

to Flossdorf and Wagner (FW)[10], Millikan (M)[8], FJM on two burners of 1.6 cm (S) and 3.2 cm (L) diameters at pressures in the range 60 to 300 torr, and ST.

In Figure 7, the data given in Reference 3 for ethane is shown. Equation 5 gives a value of $(C/O)_{neq}$ of 0.4 $(O_{neq}=1.4)$ for C_2H_6. Here too, it is observed that the flame data all lie between the $(C/O)_{neq}$ limit and the $(C/O)_{eq} = 1.0$ limit. It is interesting to note that the RA data for C_2H_6 show much less deviation from the $(C/O)_{eq} = 1$ limit than for C_2H_2 and C_2H_4. A possible explanation for the observation that relatively more oxygen is required to suppress soot formation for fuel molecules containing higher proportions of hydrogen for highly non-equilibrium conditions is offered by the recent work of Glassman, et. al. (GDC)[4] who suggest that there are three distinct but coupled zones in hydrocarbon combustion.

> (1) Following ignition, primary fuel disappears with little or no energy release and produces unsaturated hydrocarbons and hydrogen. A little of the hydrogen is concurrently being oxidized to water.
>
> (2) Subsequently, the unsaturated compounds are further oxidized to carbon monoxide and hydrogen. Simultaneously the hydrogen present and formed is oxidized to water.
>
> 3. Lastly, the large amount of carbon monoxide formed is oxidized to carbon dioxide and most of the heat release from the primary fuel is obtained.

This hydrogen combustion model of GDC is based on flow reactor experiments in which highly diluted paraffin hydrocarbon fuels were burned, primarily for very lean F/A ratios. The one set of data for a fuel-rich case which they reported indicates that their model must be modified slightly to state that the third zone effectively disappears as C/O approaches one with heat release occurring throughout the first two zones; with any CO_2 formation occurring in the second zone.

In fuel rich combustion of ethane ($\phi = 1.94$) GDC found that the C_2H_6 was very rapidly converted in the first zone to primarily C_2H_4 and H_2. Some of the excess H_2 was immediately oxidized to H_2O. The net result is that relatively more H_2O is formed in the first zone for C_2H_6 than for C_2H_2 or C_2H_4. Thus, more oxygen must be supplied to suppress soot formation in the initial zone for ethane than for C_2H_2 or C_2H_4.

The above discussion basically points to the fact that the $(C/O)_{neq}$ limit is a true lower limit for situations where the residence time is sufficient to encompass the first two zones of the (modified) GDC model of hydrocarbon combustion.

The non-equilibrium model for soot formation that we have put forward is seen to be conservative. This is due to the fact that we have assumed that all of the H_2 is immediately oxidized to H_2O. In actuality, under conditions of rich combustion, not all of the H_2 is oxidized to H_2O. The fuel-rich data of both GDC and Bonne, et. al. (BHW)[9]clearly show that when $(C/O)_{neq} < C/O < 1$, a sizeable fraction of the H_2 remains unoxidized from the very beginning of the initial zone. Thus, an improvement in the accuracy of the $(C/O)_{neq}$ sooting limit could be made by developing an expression for the ratio of H_2/H_2O which can be expected through the main oxidation region which includes the first two zones of the GDC combustion model.

Again we point out that the non-equilibrium sooting limit criteria put forth in this paper is clearly not applicable to predicting incipient soot formation during the first zone.

In Table 1, C/O and equivalence ratios corresponding to incipient soot formation for equilibrium and for non-equilibrium cases are given for a number of hydrocarbon fuels. For fuels which do not contain oxygen, it is observed that the equivalence ratio for incipient soot formation in equilibrium combustion products decreases with increasing H/C.

TABLE 1

EQUILIBRIUM AND NON-EQUILIBRIUM SOOTING LIMITS

FOR SELECTED HYDROCARBON FUELS

FAMILY	FORMULA	NAME	$(C/O)_{EQ}$	$(C/O)_{NEQ}$	ϕ_{EQ}	ϕ_{NEQ}
PARAFIN	CH_4	METHANE	1.0	.33	4.0	1.33
	C_2H_6	ETHANE	1.0	.400	3.5	1.400
	C_3H_8	PROPANE	1.0	.43	3.33	1.43
	C_4H_{10}	BUTANE	1.0	.44	3.25	1.44
	C_5H_{12}	PENTANE	1.0	.455	3.20	1.455
	C_6H_{14}	HEXANE	1.0	.462	3.167	1.462
	C_8H_{18}	OCTANE	1.0	.471	3.125	1.471
OLEFIN AND NAPTHENE	C_MH_{2M}	___YLENE CYCLO___	1.0	.50	3.0	1.50
DIOLEFIN	C_5H_8	PENTADIENE	1.0	.556	2.8	1.556
	C_6H_{10}	HEXADIENE	1.0	.545	2.833	1.545
	C_7H_{12}	HEPTADIENE	1.0	.539	2.857	1.539
CYCLOOLEFIN	C_5H_8	CYCLOPENTENE	1.0	.556	2.8	1.556
ACETYLENES	C_2H_2	ACETYLENE	1.0	.667	2.5	1.667
	C_3H_4	PROPYNE	1.0	.600	2.667	1.600
	C_4H_6	BUTYNE	1.0	.570	2.75	1.57
ALCOHOL	CH_4O	METHANOL	∞	.50	∞	1.5
	C_3H	ETHANOL	2	.5	6	1.5
	C_3H_8O	PROPANOL	1.5	.5	4.5	1.5
	$C_4H_{10}O$	BUTANOL	1.33	.5	3.0	1.5
NITRO-PARAFFIN	CH_3NO_2	NITROMETHANE	∞	2	∞	3
	$C_2H_5NO_2$	NITROETHANE	∞	.8	∞	1.8

SECTION IV

AN EQUILIBRIUM FLAME TEMPERATURE COMPUTER PROGRAM

An efficient, easy to use, and reliable computer program for calculating the equilibrium composition and flame temperature for the combustion of hydrocarbon fuels was needed by the Fuels Branch of AFAPL. Current work underway related to gas turbine engines dictated the need for a computer program which could do these calculations for air:fuel ratios ranging from slightly rich to very lean (12:1 to 100:1 for C_8H_{16} for example) while future work planned also included the need for combustion calculations for very rich mixtures up to the equilibrium sooting limit. The computer program developed to meet these needs represents a modified and improved version on an existing computer program previously developed by the author. The methods and procedures used to set up and solve the thermochemical equilibrium equations are basically those presented in NACA 1037. A brief description and "User's Manual" is given in the following paragraphs. The listing of the program (not given here) is replete with comment cards which describe the program in detail.

The following equations comprise the system.

7 chemical equilibrium equations
1 energy equation
4 mass balance equations

and there are 12 unknowns (T_f and 11 species concentrations).

$$\frac{1}{2}H_2 \rightleftarrows H \qquad K_1 = \frac{P_H}{(P_{H_2})^{1/2}} = \frac{n_H}{(n_{H_2})^{1/2}}\left(\frac{P}{n_e}\right)^{1/2}$$

$$\frac{1}{2}O_2 \rightleftarrows O \qquad K_2 = \frac{P_O}{(P_{O_2})^{1/2}} = \frac{n_O}{(n_{O_2})^{1/2}}\left(\frac{P}{n_e}\right)^{1/2}$$

$$\frac{1}{2}N_2 \rightleftarrows N \qquad K_3 = \frac{P_N}{(P_{N_2})^{1/2}} = \frac{n_N}{(n_{N_2})^{1/2}}\left(\frac{P}{n_e}\right)^{1/2}$$

$$\frac{1}{2}O_2 + \frac{1}{2}N_2 \rightleftarrows NO \qquad K_4 = \frac{P_{NO}}{(P_{O_2})^{1/2}(P_{N_2})^{1/2}} = \frac{n_{NO}}{(n_{O_2})^{1/2}(n_{N_2})^{1/2}}$$

$$\frac{1}{2} O_2 + CO \rightleftarrows CO_2 \qquad K_5 = \frac{P_{CO_2}}{P_{CO}(P_{O_2})^{1/2}} = \frac{n_{CO_2}}{n_{CO}(n_{O_2})^{1/2}} \left(\frac{P}{n_e}\right)^{1/2}$$

$$\frac{1}{2} O_2 + H_2 \rightleftarrows H_2O \qquad K_6 = \frac{P_{H_2O}}{P_{H_2}(P_{O_2})^{1/2}} = \frac{n_{H_2O}}{n_{N_2}(n_{O_2})^{1/2}} \left(\frac{P}{n_e}\right)^{-1/2}$$

$$\frac{1}{2} H_2 + \frac{1}{2} O_2 \rightleftarrows OH \qquad K_7 = \frac{P_{OH}}{(P_{O_2})^{1/2}(P_{H_2})^{1/2}} = \frac{n_{OH}}{(n_{O_2})^{1/2}(n_{H_2})^{1/2}}$$

where

$$\ln K_i = \frac{-\Delta G_i}{RT}$$

These equations can be more conveniently handled in logarithmic form i.e.,

$$0 = \ln n_H - \frac{1}{2} \ln n_{H_2} + \frac{1}{2} \ln (P/n_e) - \ln K_1 \tag{8}$$

Now a considerable simplification in the analysis occurs if P/n_e is taken to be 1.

The reasoning is as follows: The perfect gas law is

$$P/n_e = \frac{RT}{V} \tag{9}$$

then since $\frac{P_i}{n_i}$ is also equal to RT/V it follows that if RT/V is chosen to be P, then

$$P_i = n_i \tag{10}$$

and V = RT.

In other words if throughout the ensuing calculations V is always kept equal to RT, then the total amount of mass in the system must be variable so that at any temperature T, the mass of the system will always fit into a volume equal to RT. This constraint on the volume of the system is satisfied by introducing a new variable, AA which adjust the number of moles of fuel burned, NM(12), in the proper fashion such that just enough fuel is burned so that the products fit into a volume equal to RT. That is, instead of burning one mole of fuel and having the partial pressure of a species not numerically equal to the number of moles of that species, we will burn an unknown amount of fuel, AA, but we will require that AA be such that $P_i = n_i$.

Now note that only when we have guessed the correct values of the n's and T_f will the equations be satisfied. For a trial solution, we will have equation (8) not equal to 0 but equal to something we shall call δ_1 and likewise for other equations. The value of each δ_i must approach zero when the solution to the problem is found.

The complete set of chemical equilibrium equations to be solved is then

$$\left.\begin{array}{ll}
\delta_1 \ln n_H - \frac{1}{2} \ln n_{H_2} - \ln K_1 & (11) \\
\delta_2 = \ln n_O - \frac{1}{2} \ln n_{O_2} - \ln K_2 & (12) \\
\delta_3 = \ln n_N - \frac{1}{2} \ln n_{N_2} - \ln K_3 & (13) \\
\delta_4 = \ln n_{NO} - \frac{1}{2} \ln O_2 - \frac{1}{2} \ln N_2 - \ln K_4 & (14) \\
\delta_5 = \ln n_{CO_2} - \ln n_{CO} - \frac{1}{2} \ln n_{O_2} - \ln K_5 & (15) \\
\delta_6 = \ln n_{H_2O} - \ln n_{H_2} - \frac{1}{2} \ln n_{O2} - \ln K_6 & (16) \\
\delta_7 = \ln n_{OH} - \frac{1}{2} \ln n_{O_2} - \frac{1}{2} \ln n_{H_2} - \ln K_7 & (17)
\end{array}\right\} \text{Equilibrium Equations}$$

The four mass balance equations simply state that the same number of moles of any element must appear in both the products and the reactants. Thus, if AA moles of fuel are burned, then the following relations must hold:

$$\left.\begin{aligned}
\delta_8 &= AA * AO - n_H - 2n_{H_2O} - n_{OH} - 2n_{H_2} \\
\delta_9 &= AA*BO - n_O - n_{NO} - 2n_{CO_2} - n_{H_2O} - n_{OH} - 2n_{O_2} - n_{CO} \\
\delta_{10} &= AA*CO \qquad - n_N - n_{NO} - 2n_{N_2} \\
\delta_{11} &= AA*DO - n_{CO_2} - n_{CO}
\end{aligned}\right\}
\begin{aligned}
&(18)\\ &(19)\\ &(20)\\ &(21)
\end{aligned}
\ \text{Mass Balance Equations}$$

where

AO = number of moles of H in the reactants
BO = number of moles of O in the reactants
CO = number of moles of N in the reactants
DO = number of moles of C in the reactants

and where the reactants consist of one mole of fuel plus air.

Because we have set $P_i = n_i$ and thus had to introduce a new variable, AA, we must add an additional equation to form a complete *system of* 13 equations and 13 unknowns. The new equation to be added is Daltons Law of Partial Pressures which simply states that the sum of the n_i's must be equal to the total pressure. Thus:

$$\delta_{12} = P - \sum_{i=1}^{11} n_i \qquad (22) \text{ Pressure Equation}$$

Finally the energy equation is:

$$\delta_{13} = AA*HI - \sum_{i=1}^{11} n_i h_i \qquad (23) \text{ Energy Equation}$$

In the energy equation, HI is the enthalpy of the reactants for burning one mole of fuel which is multiplied by AA which is the number of moles of fuel burned.

A direct solution of Equations (11) through (23) is usually not feasible. The Newton-Raphson method for solving a system of non-linear algebraic equations is used here. A certain amount of simplicity and improved convergence can be affected by considering not the unknowns, but the logarithms of the unknowns to be the independent variables. Thus, the unknowns in the Newton-Raphson method are taken as

$$\Delta x_i = \Delta \log n_i = \log (n_i)_{r+1} - \log (n_i)_r$$

With these new definitions of the independent variables the correction equations can easily be obtained. Recall that $\frac{\partial}{\partial (\ln x)} = x \frac{\partial}{\partial_x}$ and define $q_i = \frac{\partial (\log K_i)}{\partial (\log T)}$. It can easily be shown that $\frac{\partial (\log K_i)}{\partial (\log T)} = (\frac{\Delta H^o}{RT})_i$.

Consider the first equation of the system, Equation 11:

$$-\delta_1 = \frac{\partial \delta_1}{\partial (\log n_H)} [\log (n_H)_{r+1} - \log (n_H)_r] - \frac{1}{2} \frac{\partial \delta_1}{\partial (\log n_{H_2})} [\log (n_{H_2})_{r+1} - \log (n_{H_2})_r] - q_1 [\log (T)_{r+1} - \log (T)_r]$$

and

$$\frac{\partial \delta_1}{\partial (\log n_H)} = 1 \quad \text{and} \quad \frac{\partial \delta_1}{\partial (\log n_{H_2})} = -1/2$$

The last equation, the energy equation becomes

$$\delta_{13} = AA*HI - \sum_{i=1}^{11} n_i h_i = -\sum_{i=1}^{13} \frac{\partial \delta_{13}}{\partial x_i} \Delta x_i$$

$$\delta_{13} = -AA*HI \, \Delta x_{AA} + \sum_{i=1}^{11} n_i h_i \Delta x_i + T \sum_{i=1}^{11} n_i C_{pi} \Delta x_T$$

This system of 13 equations in 13 unknowns can be expressed in the matrix form

$$-\delta_i^* = [\,C_{ij}\,]\,\Delta x_j$$

where the coefficients in the matrix $[Ci_j]$ are given in the following array

H^1	O^2	N^3	NO^4	$CO_2{}^5$	H_2O^6	OH^7	$H_2{}^8$	$O_2{}^9$	$N_2{}^{10}$	CO^{11}	AA^{12}	T^{13}
1							$-\frac{1}{2}$					$-q_1$
	1							$-\frac{1}{2}$				$-q_2$
		1							$-\frac{1}{2}$			$-q_3$
			1					$-\frac{1}{2}$	$-\frac{1}{2}$			$-q_4$
				1				$-\frac{1}{2}$		-1		$-q_5$
					1		-1	$-\frac{1}{2}$				$-q_6$
						1	$-\frac{1}{2}$	$-\frac{1}{2}$				$-q_7$
n_1					$2n_6$	n_7	$2n_8$				-AA*AO	
	n_2		n_4	$2n_5$	n_6	n_7		$2n_9$		n_{11}	-AA*BO	
		n_3	n_4						$2n_{10}$		-AA*CO	
				n_5						n_{11}	-AA*DO	
n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	n_{11}		
n_1h_1	n_2h_2	n_3h_3	n_4h_4	n_5h_5	n_6h_6	n_7h_7	n_8h_8	n_9h_9	$n_{10}h_{10}$	$n_{11}h_{11}$	-AA*HI	$T\times C'$

and where $\delta_i^* = \delta_i \qquad i = 1,\ 2,\ 3,\ \ldots\ldots 7$

$$\delta_i^* = -\delta_i' \qquad i = 8,\ \ldots\ldots 13$$

$$C' = \sum_{i=1}^{11} n_i C_{pi} \qquad i = 1,\ 2,\ 3,\ \ldots\ldots 11$$

$$q_i = \left(\frac{\Delta H^\circ}{RT}\right)_T \qquad i = 1,\ 2,\ 3,\ \ldots\ldots 11$$

In addition, the following quantities must be calculated

ΔH^o (heat of reaction) for each reaction

ΔG^o (free energy change) for each reaction

$\int_{T_d}^{T} C_{pi}\, dT$ for each species

$\int_{T_d}^{T} C_{pi} \frac{dT}{T}$ for each species

In the program, the datum temperature, Td, is taken to be 0^oR. However to avoid having to input C_p data over the entire range from 0^oR to 6000^oR the following scheme is used:

$$h = (\Delta hf^o)_{0^oR} + (h^o_{536.7^oR} - h^o_{0^oR}) + \int_{536.7^o R}^{T} cpdT$$

and the sum $(\Delta hf^o)_{0^oR} + (h^o_{536.7^oR} - h^o_{0^oR})$ is part of the input for each species.

Likewise for the entropy,

$$S^o = S^o_{536.6^oR} + \int_{536.7}^{T} \frac{C_p}{T} dT$$

and $S^o_{536.6^oR}$ is input for each species.

For each reaction, $\nu_1 A_1 + \nu_2 A_2 \rightleftarrows \nu_3 A_3 + \nu_4 A_4$

$$\Delta G^o (T) = (\nu_3 h + \nu_4 h - \nu_1 h - \nu_2 h)_T$$

and
$$\Delta G^o (T) = (\nu_3 g^o_3 + \nu_4 g^o_4 - \nu_1 g_1 - \nu_2 g^o_2)_T$$

where $g^o_i = h - Ts^o$

Constant Volume Combustion

The procedure is basically the same as for constant pressure combustion except for the following changes:

(1) Instead of taking $P_i = n_i$ and thus requiring the introduction of AA as a variable, replace $\frac{P}{n_e}$ in the equilibrium equations by $\frac{RT}{V}$. Take V (which is constant) for one mole of fuel burned. Then for instance the equilibrium equations can be written as $\nu_1 A_1 + \nu_2 A_2 \rightleftarrows \nu_3 A_3 + \nu_4 A_4$

$\delta_i = \nu_3 \ln n_3 + \nu_4 \ln n_4 - \nu_1 \ln n_1 - \nu_2 \ln n_2 + (\nu_3 + \nu_4 - \nu_1 - \nu_2) \ln \frac{RT}{V} - \ln K_i$ and $\ln \frac{RT}{V} = \ln \frac{R}{V} + \ln T$

where $\frac{R}{V}$ is constant. Thus the correction equation can be written as

$$-\delta_i = \nu_3(\Delta \ln n_3) + \nu_4(\Delta \ln n_4) - \nu_1(\Delta \ln n_1) - \nu_2(\Delta \ln n_2) - [q_i - (\nu_3 + \nu_4 - \nu_1 - \nu_2]$$
$$(\Delta \ln T)$$

Thus the only change from the constant pressure combustion case is that in column 13 of the matrix, replace q_i by $[q_i - (\nu_3 + \nu_4 - \nu_1 - \nu_2)]$ where $\nu_3 + \nu_4 - \nu_1 - \nu_2$ for each reaction is simply the sum of the first eleven matrix elements in the row corresponding to the i-th reaction

(2) Since V_i must be equal to V_f we have to introduce this equation in place of Daltons Law of Partial pressure. The proper statement of $V_i = V_f$ is to sum the partial volumes of each species i.e.,

$$\delta_{12} = V_i - \sum_i \frac{n_i RT}{P} = V_i - \frac{RT}{P} \sum n_i$$

Since P is introduced into this equation and P is unknown, we must now consider it as an independent variable. Thus instead of having AA as a variable we now have P as a variable. Thus x_{12} is now not lnAA but lnP and so in the 13x13 matrix the elements in the 12th column should be the derivatives of the 13 equations with respect to lnP. Inspection shows that all of the elements of the 12th <u>column</u> will be zero except the one in the 12th <u>row</u> since only the δ_{12} equation contains the pressure as a variable. The coefficient of $\Delta \ln P$ is from the δ_{12} equation

$$\delta_{12} = \frac{RT}{P} \sum_{i=1}^{11} n_i (\Delta \ln n_i) - \underbrace{\left(\sum_{i=1}^{11} n_i \right)}_{\text{Coefficient of } \Delta \ln P} \frac{RT}{P} \Delta \ln P + \sum_{i=1}^{11} n_i \frac{RT}{P} \Delta \ln T$$

(3) The energy equation is

$$U_i = U_f$$

instead of $H_i = H_f$

Affecting this change is elementary, i.e., $u = h - RT$ and $C_v = C_p - R$. Also note that when $\frac{P}{n_e}$ is replaced by $\frac{RT}{V}$, then $\frac{RT}{V}$ must be expressed in atmospheres whenever it appears in any of the 7 equilibrium equations. Thus if the initial pressure is expressed in atmospheres, then express R as $\frac{1545}{14.7}$ in any of the 7 equilibrium equations and also in the equation for δ_{12}.

Initial Guesses

In order to insure convergence of the program for an arbitrary fuel and arbitrary air: fuel ratios it is imperative that a foolproof method be used to generate good initial guesses for the unknowns. The following procedures for generating good initial guesses were found to work very satisfactorily. First, subroutine BAL is called which balances the primary combustion equation given the fuel composition and the air:fuel ratio. If the input A/F ratio is such that the combustion is stoichiometric or lean, complete combustion is assumed. If the input A/F ratio corresponds to slightly fuel rich combustion, (ϕ less than ϕ_{neq} given by Equation 6) then the combustion equation is balanced according to the procedure described in Section III. If the input A/F ratio lies between ϕ_{neq} and ϕ_{eq} (ϕ_{eq} given by Equation 2), the amount of CO is taken to be

$$n_{CO} = M \text{ (no. of carbon atoms in fuel) } \times \frac{\phi}{\phi_{eq}}$$

and $n_{CO_2} = M - n_{CO}$

Hydrogen and oxygen balances then determine the number of moles of H_2O and H that can be present in the primary combustion equation. Subroutine BAL thus provides a first estimate to the initial guesses on the species concentrations for CO_2, H_2O, CO, H_2, O_2, and N_2.

The second step is to obtain an estimate on the final temperature, T_f. A first estimate on T_f is obtained by solving the energy equation in the form: fuel heat of combustion = $\bar{C}_p * \sum n_i \ (T_f - T_i)$ plus heating value of CO and H_2 in products, which is solved for T_f. $\bar{C}_p$ represents the mean specific heat of the products of combustion(taken to be 9 for constant pressure combustion and 7 for constant volume combustion) and $\sum n_i$ is the total number of moles of products of combustion as determined from the output of subroutine BAL. This estimate for T_f is biased toward a lower number for ϕ 's less than 1.2 by updating T_f using the following empirical formula

$$T_f = T_f / [1.2 + \frac{2}{3} \ \phi \ (\phi - 1)].$$

The third step is to estimate the unknown, AA, for constant pressure combustion or P for constant volume combustion. Since P $/n_e$ is taken to be unity

$$\frac{P}{AA \times \sum n_i} = 1$$

or

$$AA = P \ / \sum n_i$$

for constant pressure combustion. For constant volume combustion, P_f is estimated by

$$P_f = P_i \quad \frac{\sum n_i}{n_r} \ \frac{T_f}{T_i}$$

where n_r is the number of moles of gaseous constituents in the reactants.

The fourth step is to determine the equilibrium constants at T_f. This is done in subroutine TCHEM which determines all thermochemical properties needed in the program.

The fifth step is to determine the final initial estimates for the amount of each species present in the products. This is accomplished in subroutine IGUESS which uses all of the information generated in the first four steps to come up with the guesses on the n_i. In IGUESS certain key reactions are solved independently to determine initial guesses on the major species to be expected in the products. The amount of each minor species is then estimated based on the estimates of the major species. The choice of key reactions is based on the A/F ratio and temperature. In all cases, n_{N_2} is taken to be the value of n_{N_2} determined in subroutine BAL. Also, in all cases, the minor species n_H, n_O, n_N, n_{NO}, and n_{OH} are determined by solving the individual equilibrium reaction equations (reactions 1, 2, 3, 4, and 7 respectively) in which these species occur after initial guesses on the other species involved in these reactions have been determined.

Three different procedures are used to determine the other species. If the equivalence ratio is less than 0.5 or if the equivalence ratio is less than 0.9 and the temperature T_f is less than 3000°R, the following procedure is used: n_{O_2}, n_{CO_2}, and n_{H_2O} are taken from subroutine BAL; n_{CO} is determined from the fifth reaction and n_{H_2} is determined from the sixth reaction. If the A/F is near or at stoichiometric, or rich or lean with T_f greater than 3000°R, the degree of dissociation of the fifth and sixth reactions are determined ignoring the coupling between the two reactions. This procedure provides the initial guesses for n_{CO_2}, n_{CO}, n_{H_2}, n_{H_2O}, and n_{O_2}. If ϕ is greater than 1.5 or if ϕ is greater than 1.2 and T_f is less than 3000°R, the water gas reaction

$$CO + H_2O \rightleftharpoons CO_2 + H_2$$

is used to determine initial guesses on the four species involved in this reaction. The initial guess on n_{O_2} is determined by solving the fifth reaction using n_{CO} and n_{CO_2} as determined from the water gas reaction.

These initial guesses on the n's are based on burning one mole of fuel. Unless constant volume combustion is being considered, the n's are multiplied by AA in the main program before starting the Newton-Raphson iteration procedure.

Input Card Format

Two cards are required for thermochemical input data for each of the eleven species. The data on these cards include the following: coefficients of curve fits to specific heat, c_p where

$$c_{p_i} = B(i,1,j)+B(i,2,j) * T+B(i,3,j)/T^2 \qquad j=1,2,3$$

where T is in degrees Kelvin, and c_p has units of cal/gm-mole °K, the quantity DHF_i which is defined as

$$DHF_i = \Delta h^o_{f_i}(0\ {}^oK) + h^o_i(298^oF) - h^o_i(0^oK),$$

with units of kilocalories per gram-mole, and the quantity S0 which is defined as the absolute entropy at 298°K with units of calories per gram-mole per degree Kelvin. The curve fit to c_p is made for three temperature ranges, 0-1000°K (j=1), 1000°K-3000°K(j=2), and 3000°K-4000°K (j=3). Therefore, three sets of the B's are required for each species.

The index i ranges from 1 to 11. The symbol assigned to each of the unknowns is as follows

NM(1) = n_H	NM(7) = n_{OH}	NM(13)=T_f
NM(2) = n_O	NM(8) = n_{H_2}	
NM(3) = n_N	NM(9) = n_{O_2}	
NM(4) = n_{NO}	NM(10) = n_{N_2}	
NM(5) = n_{CO_2}	NM(11) = n_{CO}	
NM(6) = n_{H_2O}	NM(12) = AA for const press. comb. = P_f for const. vol. com.	

The first data card for each species contains the following in an E 12.0 format B(i, 1, 2), B(i, 2, 2), B(i, 3, 2), B(i, 1, 3), B(i, 2, 3), B(i, 3, 3) and card number in an I8 format. The cards are numbered from 1 to 22 with the first card for each species numbered 2i - 1.

The second card of each two card set contains the following: DHF(i), S0(i), B(i, 1, 1), B(i, 3, 1) in an E12.0 format and the card number (2i) in an I8 format. As the thermochemical input data cards are read in, the cards are checked for proper sequencing of the card numbers. If they are not in the proper sequence, the program is stopped.

Three cards are required for the data necessary to carry out one run. The first data card contains the following information

The name of the fuel ---- 25 A1 format

No. of Carbon atoms in the fuel, M
No. of Hydrogen atoms in the fuel, N
No. of Oxygen atoms in the fuel, P
No. of Nitrogen atoms in the fuel, Y or Q
Air = fuel ratio (mass ratio), AF

all in F10.4 format

FLAG
= 1 for const. press comb.
= 2 for const. volume comb.
= 3 for specified initial temperature and pressure

FLAG 2
= 1 for gaseous fuel
= 2 for liquid fuel

I 2 format

The second card contains the following information (all in F15.8 format)

Heat of combustion of fuel, HTOCMB --- (btu/lbm)

Specific heat curve fit coefficients for fuel B(12, 1, 1) B(12, 2, 1), B(12, 3, 1)
where c_p has units of btu/lbm-mole - $^{\circ}$R
If the fuel is a liquid, the specific heat should be for a liquid fuel.

The third card contains the initial temperature, TI ($^{\circ}$R) and the initial pressure, PI(atmospheres) in E 10.3 format.

REFERENCES

1. Ryan, J.P., Boehman, L.I., Iden, D.J., and Preonas, D.D., Integrated Ground Tests and Analysis Program for Airborne Laser Laboratory (ALL) Low Power Windows, Section 4 "Operational Flight Environmental Load Analysis" UDRI TR-74-56, University of Dayton Research Institute, Dayton, Ohio, December 1974.

2. Boehman, L.I. and Davison, J.E., "Refractory Metals for Advanced Gas Turbine Engines for Combined Cycle Power Generation". Proceedings of Second National Conference on Energy and the Environment, Hueston Woods State Park Lodge, Ohio, Nov. 13-15, 1974, pp. 62-67.

3. Radcliffe, S.W. and Appleton, J.P., "Shock Tube Measurements of Carbon for Oxygen Atom Ratios for Incipient Soot Formation with C_2H_2 , and C_2H_6 Fuels," Fluid Mechanics Laboratory Publication No. 71-3, MIT, April 1971.

4. Glassman, I., Dryer, F.L., and Cohen, R., "Combustion of Hydrocarbons in an Adiabatic Flow Reactor: Some Considerations and Overall Correlations of Reaction Rare," Aerospace Mechanical Sciences Report No. 1223, Princeton University, April 1975.

5. Fenimore, C.P., Jones, G.W., and Moore, G.E., "Carbon Formation in Quenched Flat Flames at 1600°K, "Sixth Symposium (International) on Combustion, 1957, pp. 242-247.

6. Homann, K.H., and Wagner, H.GG.,"Untersuching des Raektionsablaufs in fetten Kohlenwasserstoff-Sauerstoff-Flammen," Ber. Bunsengesellschaft phys. Chem., 69, 165, pp. 20-35.

7. Street, J.C. and Thomas, A., "Carbon Formation in Premixed Flames", Fuel, 34, 1955, pp. 4-36.

8. Millikan, R.C., "Non-Equilibrium Soot Formation in Premixed Flames," J. Phys. Chem. 66, 1962, pp. 794-799.

9. Bonne, V., Homann, K.H., and Wagner, H.GG., "Carbon Formation in Premixed Flames," Tenth Symposium (International) on Combustion, 1065, pp. 503-512.

10. Flossdorf, J. and Wagner, H.GG., "Rubildung in Normalen und gestonten Kohlewasserstoff-Luft-Flammen," Z. Phys. Chem. N.F., 54, 1967, pp. 113-128.

11. Huff, V.N., Gordon, S., and Morrell, V.E., "General Method and Thermodynamic Tables for Computation of Equilibrium Composition and Temperature of Chemical Reactions," NACA Report 1037, 1951.

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
ELGIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

THE EFFECT OF COBALT HYDROXIDE COPRECIPITATION IN NICKEL HYDROXIDE ELECTRODES

Prepared by:	J. T. Maloy, Ph.D.
Academic Rank:	Associate Professor
Department and University:	Department of Chemistry West Virginia University
Assignment:	
(Laboratory)	Aero Propulsion
(Division)	Aerospace Power
(Branch)	Energy Conversion
USAF Research Colleague:	David F. Pickett, Ph.D.
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

EFFECT OF COBALT HYDROXIDE COPRECIPITATION IN NICKEL HYDROXIDE ELECTRODES

By

J. T. Maloy

ABSTRACT

The performance of nickel hydroxide electrodes is greatly improved by the addition of cobalt nitrate to the nickel nitrate solution used in their formation by a cathodic deposition process. The object of this investigation was the elucidation of the mechanism of cobalt hydroxide in the electrode reaction.

A nickel microelectrode was designed to study charge-discharge under potentiostatic conditions. This electrode was used in the characterization of the cathodic deposition process for nickel hydroxide using cyclic voltammetry in water-ethanol mixtures. These studies indicate that the deposition of metallic nickel along with the deposition of nickel hydroxide could be a serious problem unless the electrode potential is made sufficiently negative to cause a considerable amount of hydroxide ion generation; this problem could occur at low current densities in a constant current deposition process. These studies also indicate that the nitrite ion is not a product of the nitrate decomposition process as hitherto believed.

Nickel hydroxide films were deposited onto the microelectrode surface by the cathodic deposition process in the presence and absence of cobalt nitrate; their behavior was studied by cyclic voltammetry, chronoamperometry, and chronocoulometry.

Cyclic voltammetric studies indicate that the mode of mass and charge transport through the nickel hydroxide film is diffusion controlled in the presence and absence of cobalt hydroxide; the diffusion coefficient, however, is several orders of magnitude less than that observed in fluid solutions. High noise levels observed in the cyclic voltammograms recorded when the charged material was present on the electrode surface indicate that the film impedance increases significantly in the charged state; this suggests that the battery active material behaves like a semiconductor in which the conductivity depends upon the state of charge. Most importantly, cyclic voltammetry reveals that cobalt hydroxide coprecipitation increases the reversibility of electron transfer in the charge-discharge cycle; in the absence of cobalt hydroxide a difference of 150 mV is observed between charge and discharge peak potentials; in the presence of ca. 10% cobalt hydroxide, this difference is only 75 mV.

Chronoamperometric studies at various fixed charging potentials reveal that the coprecipitation of cobalt hydroxide increases the rate of charge significantly. The magnitude of the charge accepted by the microelectrode is increased in the presence of cobalt hydroxide, also; this is probably due to thicker film formation when cobalt is present.

Chronocoulometric studies show that the microelectrode may be fully charged under potentiostatic conditions within 500 sec. The potential regime of efficient charge acceptance is 0.10 volt broader in the presence of cobalt hydroxide. The efficiency of charge recovery approaches 90% in this potential regime in the presence of cobalt hydroxide; in its absence, the maximum recovery observed was less than 60%, presumably due to the parasitic oxidation of solvent at the potentials sufficiently positive to induce charge in the absence of cobalt. From the magnitude of charge recovered, the thickness of the nickel hydroxide film is *ca*. 2.5μ in the presence of cobalt. In its absence, the film is believed to be no more than 200 mμ thick.

INTRODUCTION

Positive nickel hydroxide electrodes are prepared on a pilot plant scale by the Air Force Aero-Propulsion Laboratory's Battery Group at Wright-Patterson AFB, Ohio using a cathodic deposition process developed in-house (1, 2). In this process, sintered nickel plates are immersed in a water-ethanol solution of nickel nitrate at temperatures nearing the boiling point of the solution and cathodized at a constant current density (ca. 0.35 amp/in^2) for a predetermined period of time (ca. 100 min). The current density selected is sufficient to cause the solvent-supporting electrolyte system to decompose to liberate hydroxide ion. This may occur upon the direct liberation of hydrogen gas at the electrode

$$2H_2O + 2e^- \longrightarrow H_2 + 2OH^- \qquad (1)$$

or by the reduction of nitrate ion

$$\begin{array}{l} H_2O + NO_3^- + 2e^- \longrightarrow NO_2^- + 2OH^- \\ \quad \vdots \qquad \vdots \qquad \vdots \qquad\quad \vdots \qquad \vdots \\ 6H_2O + NO_3^- + 8e^- \longrightarrow NH_3 + 9OH^- \end{array} \qquad (2)$$

where a variety of reduction products may form in addition to OH$^-$. This hydroxide ion then diffuses away from the electrode and reacts with nickel ion present in the bulk of the solution to form the battery active nickel hydroxide within the pores of the nickel sinter:

$$Ni^{++} + 2OH^- \longrightarrow Ni(OH)_2 \qquad (3)$$

The deposition of metallic nickel within the pores of the sinter (which can occur at the potentials employed--see discussion below)

$$Ni^{++} + 2e \longrightarrow Ni \qquad (4)$$

is presumably prevented by the ongoing flux of OH$^-$ away from the electrode; thus, reaction 3 occurs within a finite reaction zone removed from the electrode surface in a manner not unlike that encountered in the study of electrogenerated chemiluminescence (3,4).

This positive plate (now impregnated with $Ni(OH)_2$) may be washed to remove any nitrate ion or undesirable reduction products and immersed in 30% aqueous KOH and charged. In the charging process the $Ni(OH)_2$--which exists in α and β crystaline forms--is converted to NiOOH

$$Ni(OH)_2 \longrightarrow NiOOH + H^+ + e^- \qquad (5)$$

which remains on the electrode surface (5). It has been proposed that the rate of proton diffusion out of the hydroxide film governs the rate of the charging reaction, but the agreement between the theoretical consequences of this proposal and experimental observations leave much to be desired (6).

Alternately, one could view the charging reaction to be governed by the rate of hydroxide ion diffusion into the nickel hydroxide film

$$OH^- + Ni(OH)_2 \longrightarrow NiOOH + H_2O + e^- \qquad (6)$$

Discharge may be viewed as the reverse of reaction 5 or 6. Reaction 6 has some advantage in this respect. In basic solution one would not expect large quantities of hydrogen ion to be present as is required by the reverse of reaction 5. Sufficient H_2O is present even in basic solution, however, to allow discharge to occur through the reverse of reaction 6.

OBJECTIVES

Small amounts (up to 15%) of cobalt nitrate are sometimes added to the nickel nitrate used in the cathodic deposition bath. This results in the coprecipitation of $Co(OH)_2$ along with $Ni(OH)_2$ in a reaction similar to that of equation 3. Electrodes prepared in this manner exhibit improved charging efficiency and increased cycle-life performance. The mechanism of the cobalt hydroxide participation in the electrode behavior is not understood. It is the object of this report to investigate this participation so that the mechanism may be more fully understood in order that the optimum amount of cobalt coprecipitation may be predicted and the feasibility of other additives may be investigated in a systematic manner.

EXPERIMENTAL

Cyclic voltammetry, chronoamperometry and chronocoulometry experiments were conducted with a Princeton Applied Research Model 170 Electrochemical System (PAR). The PAR was also used in the cathodic deposition process.

To study the electrode behavior under potentiostatic conditions, a nickel hydroxide microelectrode was designed and constructed. This was necessary to prevent the observation of mixed electrode potentials as would be expected in large (battery size) plates. To accomplish this, a 0.05 in. diameter nickel wire was forced-fit through a teflon sleeve and the exposed end area was ground flat. The resulting planar disk was etched in 6M HCl and then subjected to potentiostatic cathodic deposition of $Ni(OH)_2$; typically, the potential of deposition was -1.5 volts vs. SCE.

The SCE reference electrode was also manufactured in-house. A small pinhole was allowed to remain in the sidearm of the electrode that extended into the solution. This allowed this reference to be used in highly alkaline solutions. Following use, the hydroxide containminated chloride solution contained in the sidearm was allowed to drain from the electrode through this pinhole; this was then replaced with fresh saturated KCl solution. Thus, all potentials cited herein are against SCE, even in the case of alkaline solution potentials.

After the microelectrode had been subjected to the cathodic deposition process developed on a pilot plant basis at WPAFB, the Aero-Propulsion Laboratory, it was transferred to a 30% KOH solution and used in charge-discharge studies via the electrochemical techniques cited above. All charge-discharge studies were conducted at room temperature; cathodic deposition was conducted at the boiling point of the water-ethanol mixture as done routinely in the AF Aero Propulsion Laboratory (AFAPL) process. Deposition of the nickel hydroxide film occurred from 1.8 M $Ni(NO_3)_2$ solution in the absence of $Co(NO_3)_2$. Following experimentation with this electrode, the battery active material was ground away and then replaced with a fresh $Ni(OH)_2$ film deposited from a solution containing 0.18 M $Co(NO_3)_2$. The comparison of these two electrodes constitutes the bulk of this report. Some cyclic voltammetry experiments were performed on the water-ethanol deposition bath using the bare nickel electrode. These are discussed immediately below.

RESULTS AND DISCUSSION

Some information was sought concerning the cathodic deposition process. The role of the nitrate ion was investigated by carrying out the voltammetry on saturated KNO_3 in the absence of nickel. This is illustrated in Figure 1.a. Note that a sweep into the anodic background results in the generation of nickel ion; this is electrodeposited in the vicinity of - 1.0 volt vs. SCE in sweep 2 of curve a. The appearance of this reduction peak indicates that nickel deposition can occur at potentials positive of the cathodic limit (where hydroxide ion is generated). Thus, some care must be exercised in the cathodic deposition process to adjust the electrode potential sufficiently negative (at moderate current density) to generate enough hydroxide ion to shield the electrode from nickel electrodeposition. Battery plates fabricated at low current densities in preliminary experiments at AFAPL appeared gray in color and yielded poor charging characteristics (7). These voltammetric results suggest that this was probably due to nickel electrodeposition at electrode potentials too positive for hydroxide generation.

Curve 1.b illustrates the effect of adding small amounts of nickel ion to the saturated KNO_3 solution. The large oxidation wave in the vicinity of 0.4 volts on the second scan is due to the charging of $Ni(OH)_2$ formed on the electrode surface during excursions into the limitnig processes. The corresponding discharge occurs at 0.05 volts at the conclusion of the first and second cycles, indicating that the charged battery active material is partially formed during the nickel oxidation that occurs in the first cycle. Since these experiments were conducted at room temperature, these results suggest that the formation of battery active material can be accomplished at reduced temperatures. Boiling point temperatures are maintained in the AFAPL process to assure proper impregnation of the nickel sinter, however.

An investigation of the cathodic background is illustrated in Figure 2. These voltammograms show the appearance of an oxidation wave in the vicinity of 0.6 volts following an excursion into cathodic background. This is probably due to some product that is generated along with hydroxide ion in the reduction of nitrate ion. The corresponding voltammetry in the presence of nickel ion is quite similar. This indicates that the peak at 0.6 volts is due to the nitrate reduction and not to nickel participation. It has been proposed that nitrite formation during nitrate reduction causes the oxidation peak at 0.6 volts (8). This hypothesis is investigated in Figure 3 which illustrates the oxidation wave observed in detail. This shows clearly that nitrite oxidation (curve c) occurs 0.3 volts positive of the process in question.

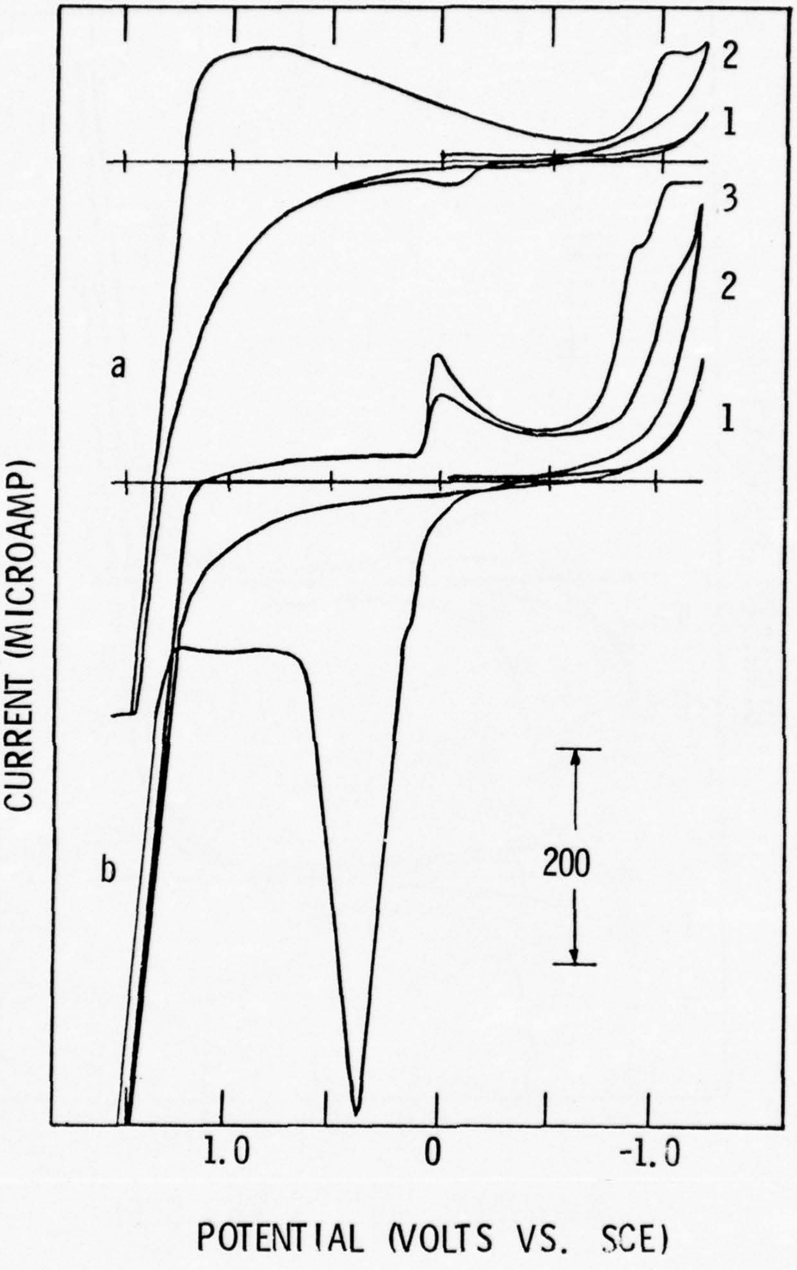


FIGURE 1: Multiple scan voltammetry into the anodic background process for water-ethanol mixtures saturated with KNO_3. A bare nickel electrode was employed and the scan rate was 1 volt/sec. The numbers designate the cycle number. Curve a: saturated KNO_3. Curve b: saturated KNO_3 containing 2 g/L $Ni(NO_3)_2$.

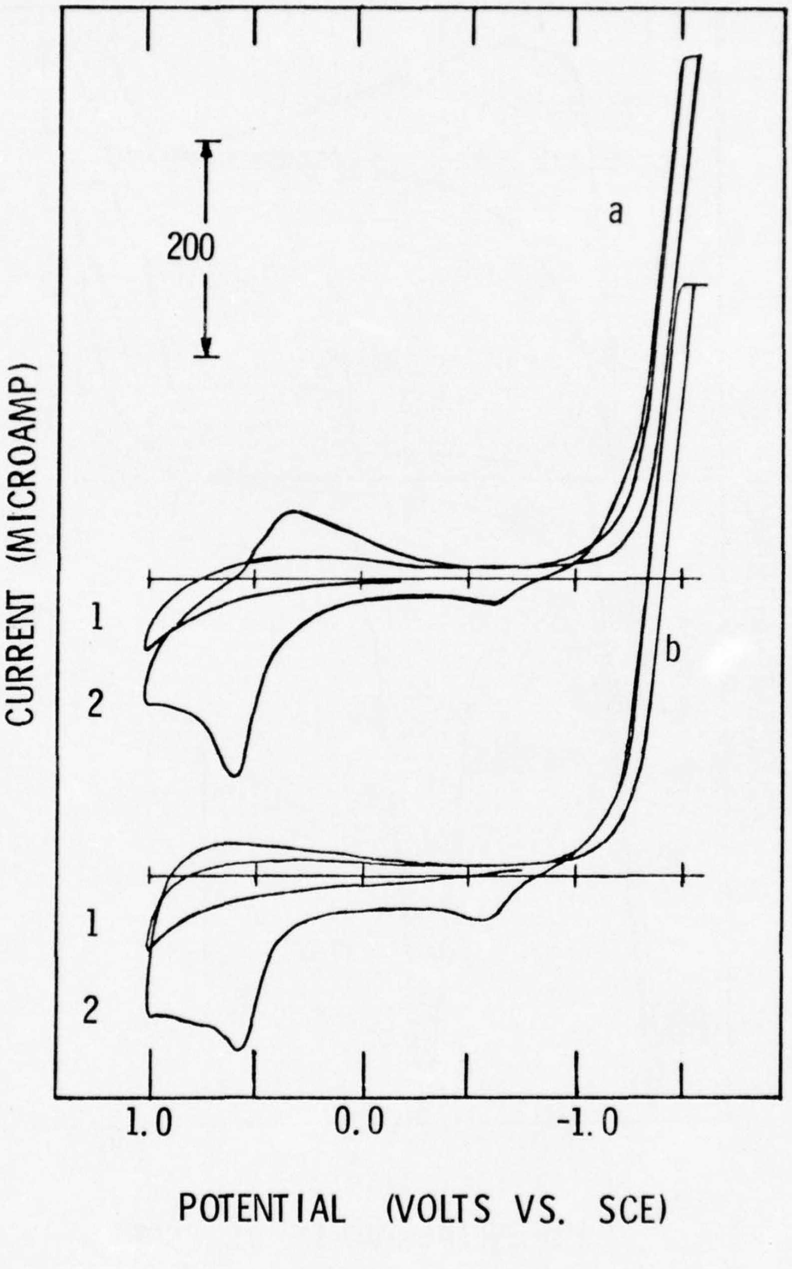


FIGURE 2: Multiple scan voltammetry into the cathodic background process for water-ethanol mixtures satuarated with KNO_3. Consult Figure 1 for experimental details.

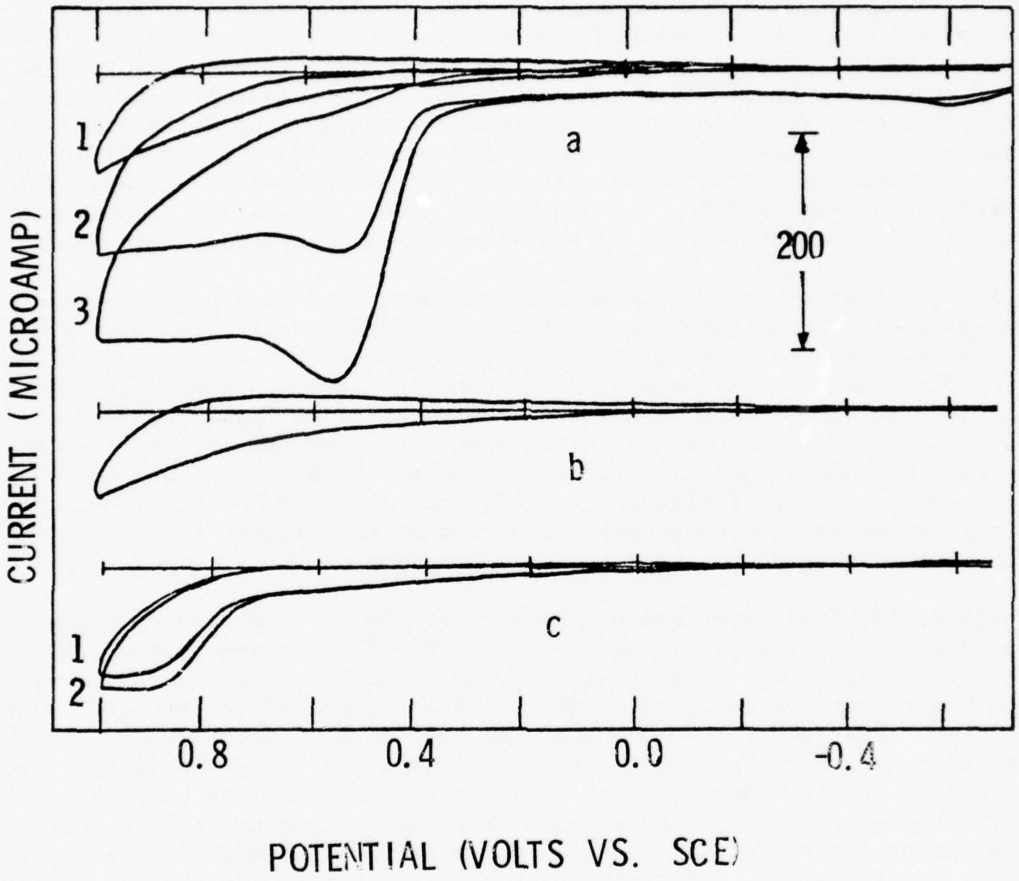


FIGURE 3: An investigation of the products of nitrate reduction. Curve a, sweep 1 and curve b: background sweeps in saturated KNO_3 without reaching cathodic limit. Curve a, sweeps 2 and 3: the oxidation of the product of nitrate reduciton during the cathodic background limiting process. Curve c: scans over the same range as curve b with 0.010 M $NaNO_2$ added. The scan rate was 0.5 volt/sec throughout.

Thus, it is unlikely that the reduction of nitrate to nitrite accounts for the hydroxide formed upon cathodization (see Equation 2). Other nitrate reduction products could correspond to this wave, however, and some investigation is necessary to determine which, if any, is.

These studies of the deposition process also revealed that potentials too negative (current densities too high) had deleterious effects. When very negative potentials were employed (in the vicinity of -2.0 volts) the green (hydrated) nickel hydroxide formed on the electrode. Apparently, the generation of too much hydroxide ion causes the $Ni(OH)_2$ precipitation to occur at greater distances from the metallic electrode. Thus, deposition occurs in a water-rich environment and the hydrated form of $Ni(OH)_2$ is deposited on the plate. This suggests that potentiostatic control may result in ideal $Ni(OH)_2$ deposition, thereby avoiding the problem of nickel deposition at potentials too positive or the deposition of the hydrated form at potentials too negative.

Electrodes prepared in this manner were subjected to slow scan rate cyclic voltammetry studies in alkaline solution. Typical results are shown in Figure 4 where the behavior of the electrode in the absence of coprecipitated cobalt hydroxide is contrasted with the behavior in its presence (curves b, c, and d). Extremely slow scan rates were employed in these studies; if rates much higher than 1 millivolt/sec were used, current maxima were not observed. The appearance of current maxima as obtained in Figure 4 indicate that the process is diffusion controlled; that these can be obtained only at slow scan rates may indicate that the diffusion coefficient for the process is much less than that observed in fluid solution.

Regardless of the presence or absence of cobalt, these voltammograms exhibit a considerable increase of recorder noise when the charging process takes place. This noise, which is apparently due to 60 hz pickup, persists as long as the battery active material remains in the charged state, but decreases upon discharge. This result suggests that the impedance of the $Ni(OH)_2$ film increases appreciably in the charged state, thereby increasing the probability of 60 hz noise pickup. This hypothesis was verified by measuring the cell resistance with the microelectrode in the charged state and comparing it with that obtained in the discharged state. A six-fold increase of cell resistance was detected when the electrode was held in the charged state. This observation, combined with that immediately above, leads one to the conclusion that the primary mode of charge transport through a $Ni(OH)_2$ electrode is via diffusion through a solid in which the conductivity depends upon the state of charge. This behavior suggests that the battery active material has semiconductor properties. It also suggests that the measurement of plate resistance may give some indication of the state of charge, but no experimentation along these lines has been attempted.

Insofar as the voltammetry is concerned, the coprecipitation of cobalt hydroxide has an obvious effect: it increases the reversibility of the electron transfer process. In its absence (Figure 4.a), the primary charging process occurs at +0.23 volts; a shoulder at 0.18 volts is probably due to the separate oxidation of α-$Ni(OH)_2$ (9). Discharge occurs at +0.08 volts. This 150 millivolt peak separation is indicative of a quasi-reversible electron transfer process in fluid solution voltammetry. The effect of cobalt coprecipitation is shown in the remaining curves. The charging peak now appears at +0.125 volts while the discharge maximum occurs at 0.045 volts. This 80 millivolt

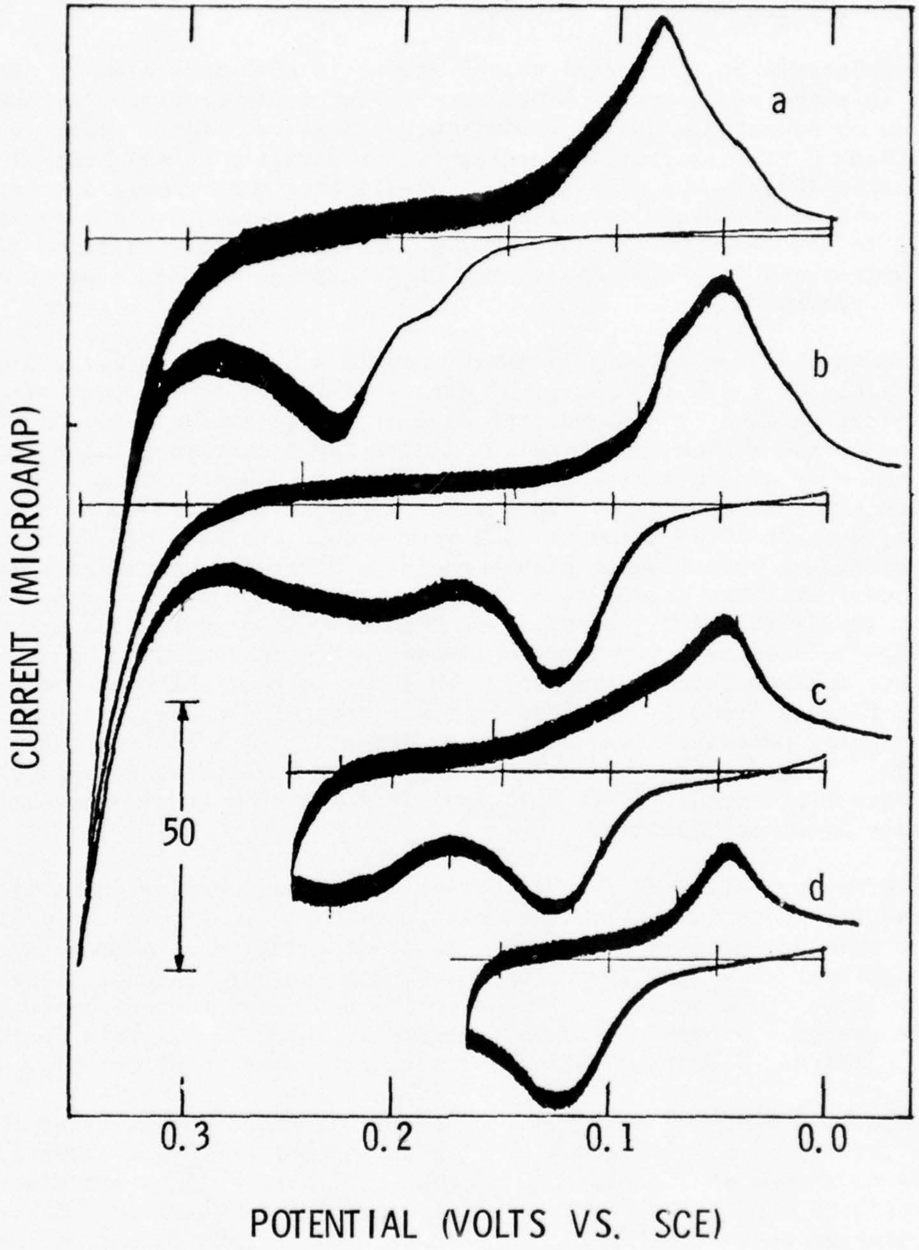


FIGURE 4: Slow scan cyclic voltammetry of 30% KOH solutions at the nickel hydroxide electrode. Curve a: containing no coprecipitated $Co(OH)_2$. Curves b, c, and d: containing $Co(OH)_2$ coprecipitated from 1.8 M $Ni(NO_3)_2$ containing 0.18 M $Co(NO_3)_2$. The scan rate employed was 0.5 millivolt/sec. The solutions were stirred during the recording of these scans.

separation lies much closer to the 60 millivolt minimum separation associated with complete reversibility in fluid solution. Thus, cobalt coprecipitation lowers the applied potential necessary for charging.

The existence of additional charge states is also indicated in curves c and d. In curve c one may note two separate charging processes occurring, even when no measurable oxygen evolution takes place. Close comparison of curves c and d reveals that these discharge separately at slightly different potentials. (Comparison with curve a reveals that the primary discharge peak in the presence of cobalt corresponds to a shoulder on the discharge peak recorded in its absence.) Thus, greater than 100% battery charging efficiencies obtained in the presence of cobalt may be due to the existance of this additional oxidation process.

Chronoamperometry at $Ni(OH)_2$ electrodes is shown in Figures 5, 6 , and 7. At potentials at which the charging current dominates, the effect of cobalt coprecipitation clearly enhances the rate of charge acceptance. In Figure 5, curves a, b, and c show the effect of double layer charging; the charging spike occurring at the instant of potential step is due to this. It becomes insignificant with respect to the initial faradaic current when the charging potential exceeds +0.28 volts vs. SCE (curve c). Curves e and f show the effect of oxygen evolution at highly positive electrode potentials; this initial current spike is due to a faradaic process (oxygen evolution) but it does not result in $Ni(OH)_2$ charging. At intermediate potentials the charging of $Ni(OH)_2$ is seen as a slow process, sometimes requiring as much as 5 seconds to achieve maximum rate (curve 5.b). This may be contrasted with the behavior shown in Figures 6 and 7, obtained in the presence of coprecipitated $Co(OH)_2$. At no charging potential is a delay time necessary to achieve a maximum rate of charge. In addition, the current-time curves in Figures 6 and 7 exhibit a much more rapid decay. This also indicates the acceleration of charge acceptance by cobalt addition.

These Figures also verify the current potential behavior shown in Figure 4. Note that the potentials of charge acceptance shown in Figure 6 are all more negative than any shown in Figure 5. As predicted by the voltammetry, charge acceptance occurs at much lower potentials when $Co(OH)_2$ is coprecipitated with $Ni(OH)_2$. This could account for some of the increased current observed at the same charging potential (compare Figure 5.a with 7.a). This could also be due to thicker $Ni(OH)_2$ film formation in the presence of $Co(OH)_2$.

Chronocoulometry was also used to define the rate and extent of charge. Typical results of double potential step chronocoulometry are shown in Figure 8. Note that even in the presence of coprecipitated $Co(OH)_2$ and under potentiostatic conditions, the charge delivered does not reach a maximum. Upon discharge to the rest potential, however, a maximum amount of charge recovery is observed. This indicates that the efficiency of charge acceptance (as measured by the ratio of charge recovered to that delivered) may decrease with increasing potentiostatic charging time.

This is verified by the data shown in Figure 9. This indicates that an on-going parasitic reaction occurs constantly even under these conditions when very little oxygen evolution is believed to occur. This could reduce the charging efficiency in $Ni(OH)_2$ plates. This also indicates that little

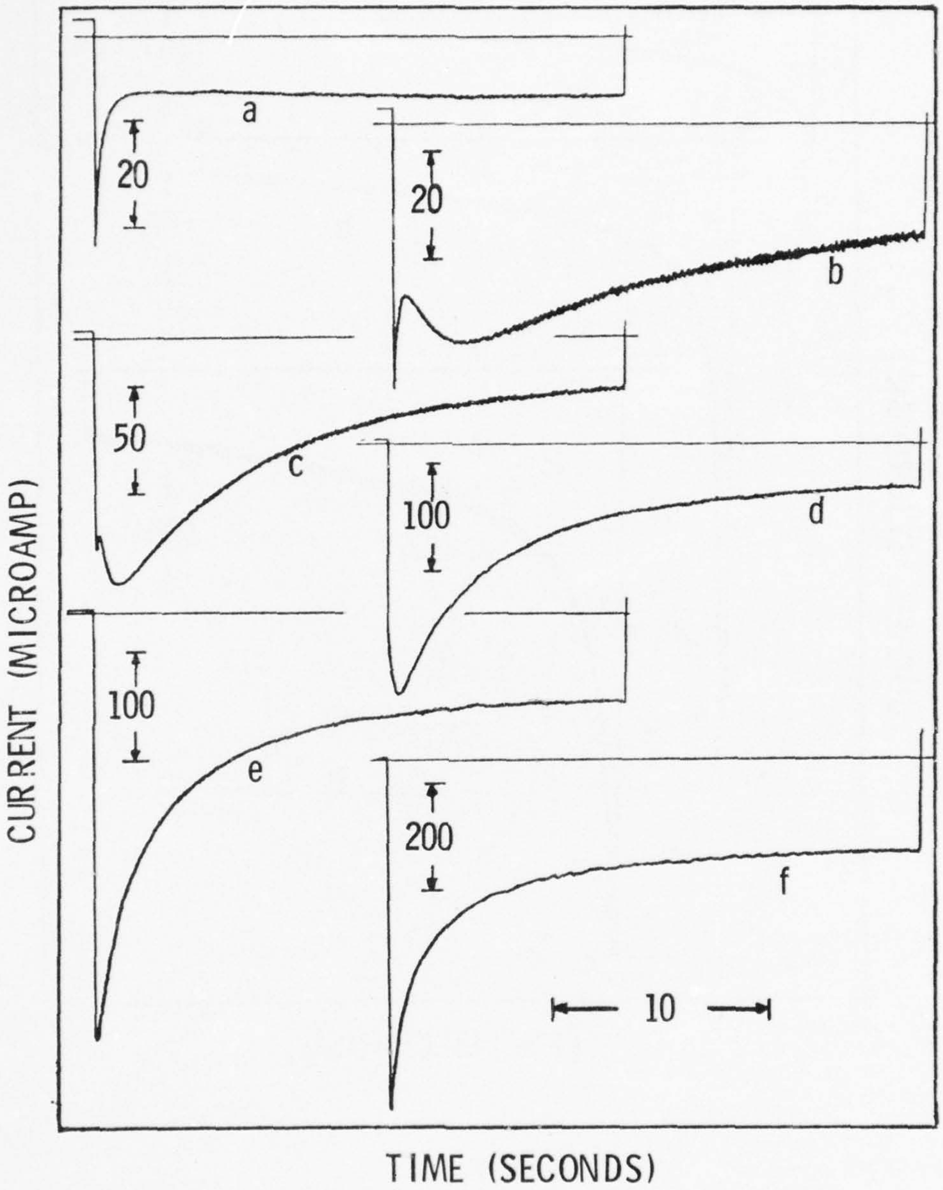


FIGURE 5: Single potential step chronoamperometry of 30% KOH at a cobalt-free $Ni(OH)_2$ electrode. Curves a through f were obtained by stepping to 0.24, 0.26, 0.28, 0.30, 0.32, and 0.34 volts vs SCE respectively to bring about the charging process.

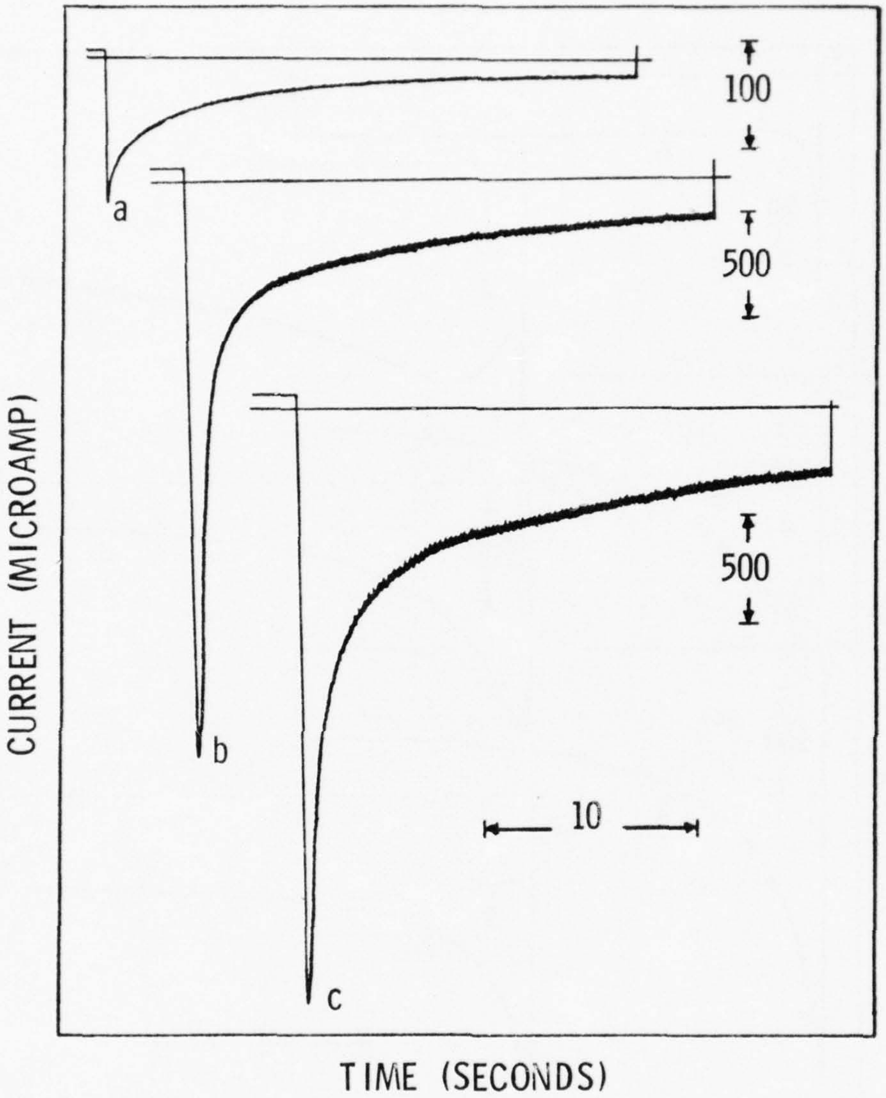


FIGURE 6: Single step chronoamperometry of KOH solution at a $Ni(OH)_2$ electrode containing ca. 10% $Co(OH)_2$. Potential steps to 0.10, 0.15, and 0.20 volts vs SCE were used in curves a, b, and c respectively.

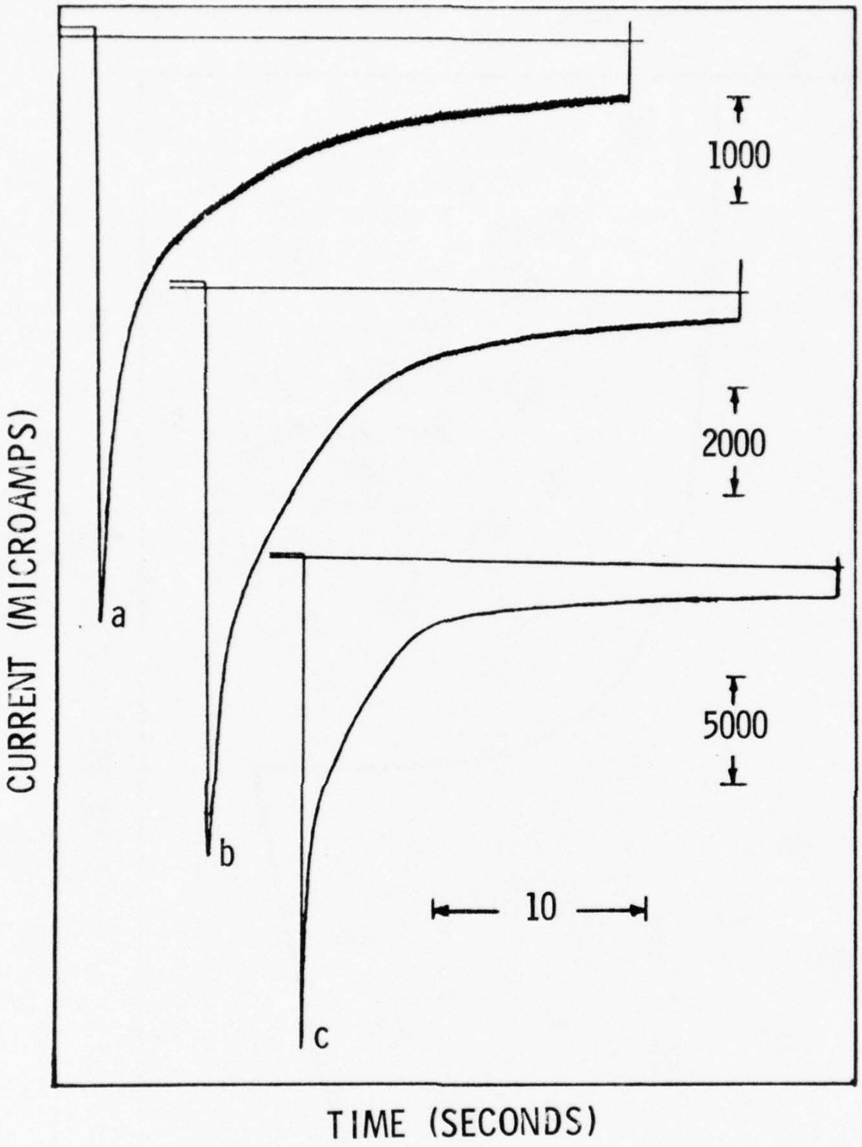


FIGURE 7: Single step chrononamperometry of KOH solution at a $Ni(OH)_2$ electrode containing $Co(OH)_2$ (continued). Potential steps to 0.24, 0.28, and 0.32 were used in curves a, b, and c respectively.

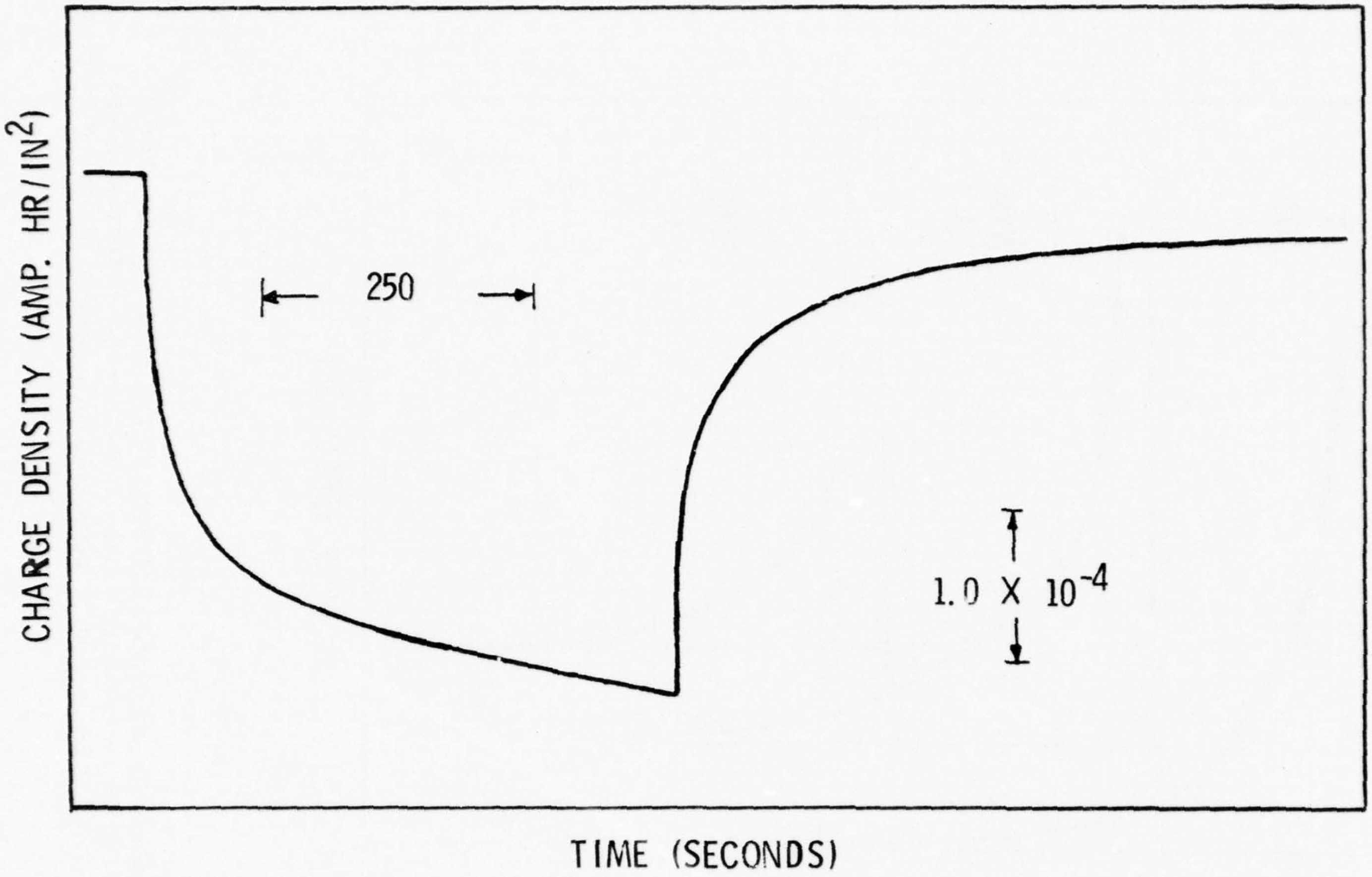


FIGURE 8: Double potential step chronocoulometry using a $Ni(OH)_2$ electrode containing $Co(OH)_2$. The potential was stepped from the rest potential (zero current) to +0.175 volts vs SCE and then back to the rest potential after a 500 sec charge.

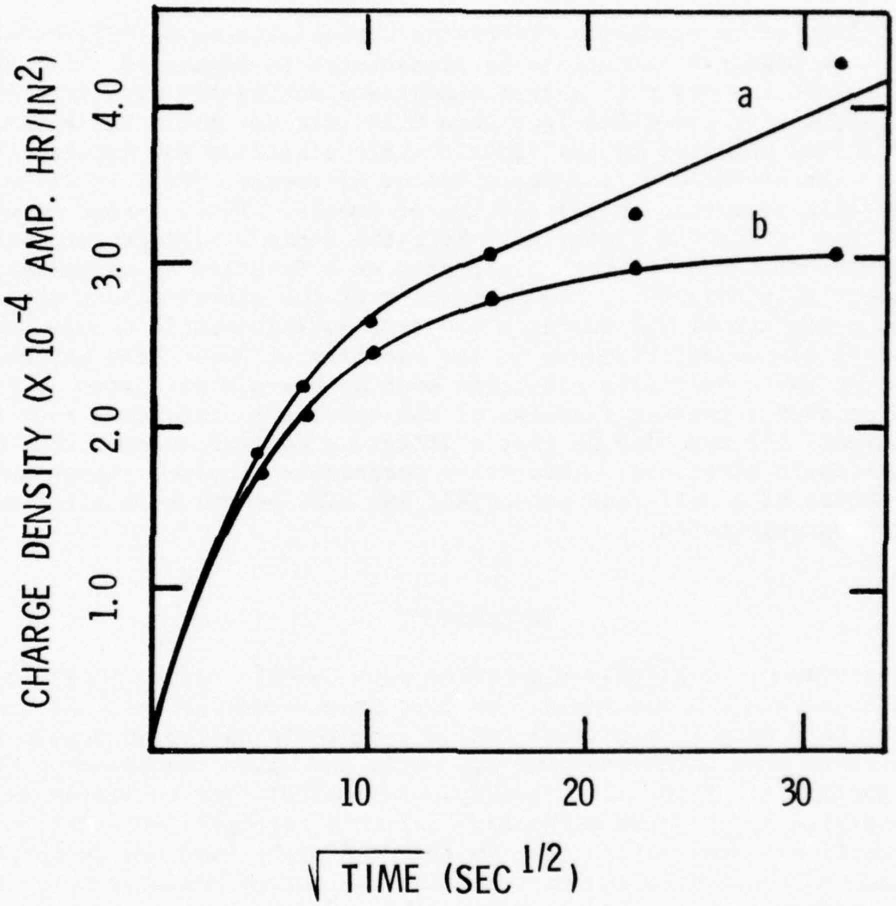


FIGURE 9: A comparison of charge recovered to that delivered under potentiostatic conditions at various charging times. Curve a: delivered. Curve b: recovered. Other experimental conditions are given in Figure 8.

is to be gained through potentiostatic charging for more than 500 seconds with electrodes of this type; insignificant additional recoverable charge was obtained upon 1000 seconds of charge. Thus, subsequent experiments were conducted using a 500 sec. charge. Of course, data of the form of curve b allows one to estimate the thickness of the $Ni(OH)_2$ film using Faraday's law and the known density of $Ni(OH)_2$ (2.5 g/cm^3). From the result of this calculation, the $Ni(OH)_2$ film, in this case, appears to have been 2.5μ thick.

The ability of an electrode containing coprecipitated $Co(OH)_2$ to accept charge at more negative potentials is illustrated in Figure 10. For the nickel-cobalt electrode the onset of charge acceptance during 500 sec. of electrolysis occurs at a positive potential less than 0.10 volt vs. SCE. The maximum amount of charge accepted by the nickel-cobalt electrode was greater than that accepted by the nickel electrode by a factor of twenty. This could be due to thinner film formation in the absence of cobalt. Thus, a more meaningful comparison may be made in Figure 11. Here the ratio of charge recovered to that delivered is shown for both electrodes as a function of charging potential. As is evident from this plot, the efficiency of the nickel-cobalt electrode approaches a maximum of 90% during a 500 sec. potentiostatic charge, while that of the nickel electrode maximizes in the vicinity of 50%. This may be due to the fact that the cobalt-free electrode must be charged at a more positive potential so that a greater fraction of the current in this case goes to produce oxygen. It may also be that a different species accepts the charge in the nickel-cobalt electrode. This seems reasonable, because charge acceptance not only occurs at a different potential, but also occurs more efficiently if $Co(OH)_2$ is coprecipitated.

CONCLUSIONS

The improvement of $Ni(OH)_2$ electrodes with $Co(OH)_2$ coprecipitation has been clearly demonstrated. Some insight to this improvement process has been provided by this work. In general, $Co(OH)_2$ coprecipitation increases the potential range over which charging may occur and makes the electron transfer process reversible. Thus, the coprecipitated $Co(OH)_2$ may be viewed as an electrocatalytic agent. The mechanism of this catalysis is still unknown. Careful scruting of the voltammetry in Figure 4 might lead one to propose that the presence of the cobalt causes the principle charge accepter to be the α-$Ni(OH)_2$ instead of β-$Ni(OH)_2$ as in the absence of cobalt, but no hard evidence exists to indicate this. If this is the case, though, the improvement in reversibility may be due to the fact that α-$Ni(OH)_2$ undergoes charge-discharge more reversibly than β-$Ni(OH)_2$.

Evidence that potentiostatic charging is less than 100% efficient even in the absence of oxygen evolution is most interesting. This aspect of the problem should be investigated in more detail because it is a fundamental problem in energy conversion.

The most promising direction for future fundamental research is in the area of charge and mass transport in solid film electrodes of this sort. This problem is not unlike a study of immobilized enzyme electrodes recently reported (10) in that diffusion and chemical reaction take place in a film deposited on a metallic electrode. This situation may be modeled using digital simulation techniques--closed form solutions to the differential equations describing

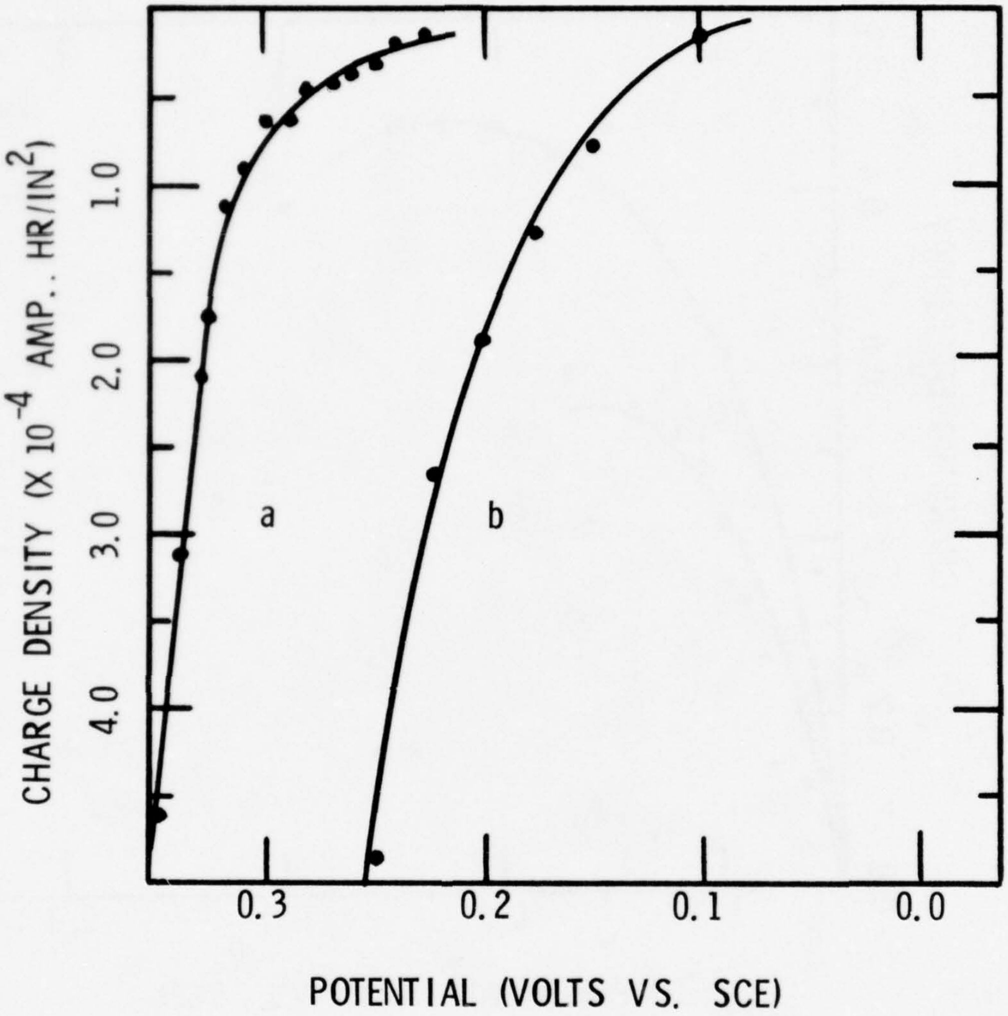


FIGURE 10: A comparison of charge delivered as a function of electrode potential in the presence and absence of coprecipitated $Co(OH)_2$. A 500 sec charge was employed throughout. Curve a: cobalt absent. Curve b: present.

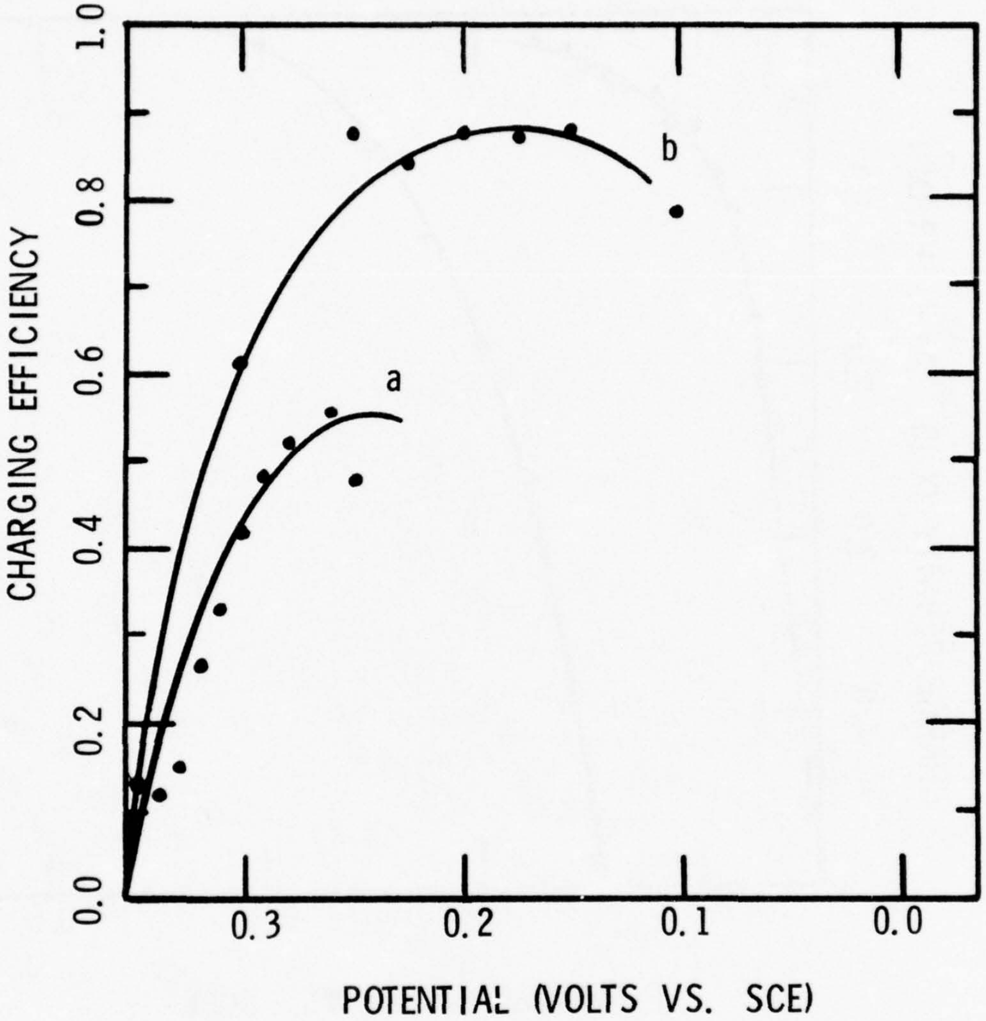


FIGURE 11: Charging efficiency as a function of electrode potential. Experimental conditions were identical to those given in Figure 10.

processes of this sort are usually difficult to obtain--with the hope of being able to predict charge-discharge characteristics on the basis of more fundamental variables (cobalt concentration, diffusion coefficients, film thickness, etc.). Whatever model is developed, it would have to describe the current-time data in Figures 5, 6, and 7 in addition to the voltammetry in Figure 4.

Finally, more experimental work is needed to fully characterize the system. Due to time limitations, only one electrode containing coprecipitated $Co(OH)_2$ was used during the course of this study. Some work is necessary to determine the effect of variations in cobalt concentration on the basic experiments outlined above. In addition, all this work was carried out at room temperature. Since real batteries must operate at low and high temperatures, much more must be known about electrode performance at temperature extremes. The experiments reported above do indicate an approach that may be taken in a study of this type. Much useful scientific information can be gained from their continuation.

REFERENCES

1. D.F. Pickett, Extended Abstracts of 144th Electrochem. Soc. Mtg, Oct. 1973, Boston, Abstract No. 49.

2. D.F. Pickett, U.S. Patent No. 3827911.

3. J.T. Maloy and A.J. Bard, J. Amer. Chem. Soc., 93, 5959 (1971).

4. T.M. Huret and J.T. Maloy, J. Electrochem. Soc., 121, 1178 (1974).

5. S.U. Falk and A.J. Salkind, Alkaline Storage Batteries, John Wiley and Sons, New York, 1969.

6. D.M. MacArthur, J. Electrochem. Soc., 117, 729 (1970).

7. J.W. Logsdon, personal communication.

8. D.M. MacArthur, "Electrochemical Properties of Nickel Hydroxide Electrodes", Power Sources, 3, D.H. Collins, ed., Oriel Press, Newcastle upon Tyne, England, 1971.

9. D.M. MacArthur, J. Electrochem. Soc., 117, 422 (1970).

10. L. D. Mell and J.T. Maloy, Anal. Chem., 47, 299 (1975).

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO

(CONDUCTED BY AUBURN UNIVERSITY)

OVERLOAD PROTECTION AND FILTERING REQUIREMENTS FOR PHASE CONTROL VOLTAGE REGULATORS

Prepared by:	Thomas A. Stuart
Academic Rank:	Assistant Professor
Department and University:	Electrical Engineering Dept. University of Toledo
Assignment:	
(Laboratory)	Aero Propulsion Laboratory
(Division)	Aerospace Power Division
(Branch)	Power Distribution Branch
USAF Research Colleagues:	P.C. Herren and P.E. Stover
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

OVERLOAD PROTECTION AND FILTERING REQUIREMENTS
FOR
PHASE CONTROL VOLTAGE REGULATORS

by

Thomas A. Stuart

ABSTRACT

The recent development of high voltage superconducting generators and freon cooled transformers has created a need for phase controlled voltage regulators operating in the 60 K.V. range. These regulators are intended for airborne missions and must be much lighter in weight than present state-of-the-art equipment. This weight requirement creates special design problems for these systems and usually requires that some compromise be made between certain specifications.

One problem of particular interest is that of minimizing the weight of the LC output filter without unduly sacrificing reliability or performance. To accomplish this task it will be necessary to (1) employ some type of current limiting circuit that will allow the use of a small series inductance (L), and (2) develop a systematic method of minimizing filter weight.

To demonstrate the feasibility of item (1), a current limiting circuit was designed and tested in the laboratory. This circuit operated successfully at the low power levels for which it was designed (50 watts), and it should be possible to extend this design to much higher power levels. The approach to item (2) was to develop a computer program which will minimize the weight of the LC filter, subject to certain design constraints. This program appears to be quite useful for evaluating certain engineering trade-offs, and it allows the variation of a wide range of design parameters.

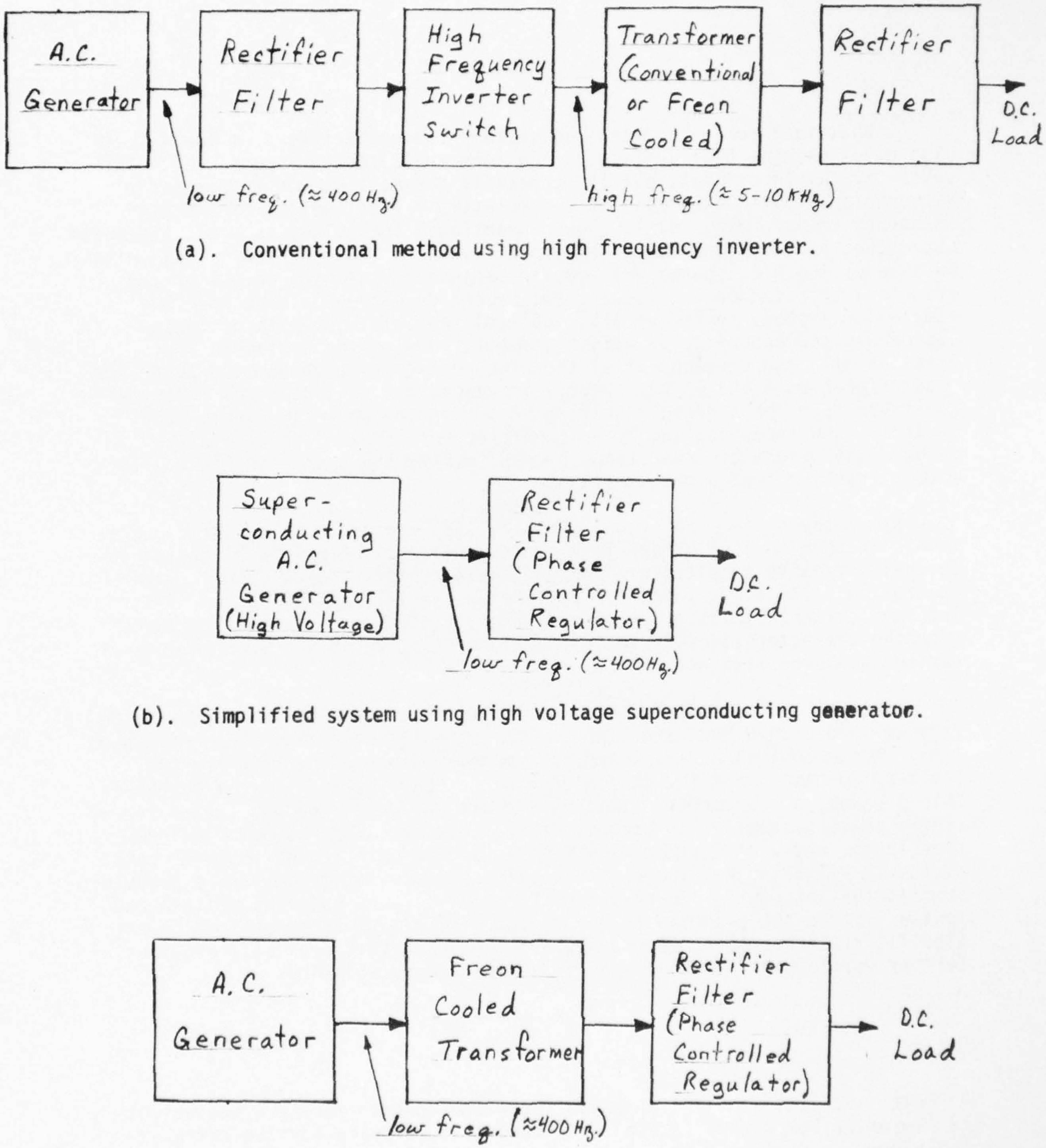
I. INTRODUCTION

Lightweight power conditioning systems for converting from low A.C. to high D.C. voltages usually take on the form indicated in Figure 1(a). This system is fairly complex, but it represents the most common approach for airborne applications because of the relatively light weight of the high frequency transformer. As long as conventional technology is used, alternate approaches have not proven to be competitive with that in Figure 1(a), either because of low A.C. source voltages or because of the high weight of low frequency transformers. However, two recent developments have made the simplified systems in Figure 1(b) and 1(c) feasible, and both of these approaches now appear to be weight competitive with the system in 1(a). First of all, the development of the high voltage superconducting generator (see references 5 and 6) has greatly increased the voltage level directly available from the source, resulting in the system shown in Figure 1(b). Secondly, new freon cooling techniques (see reference 7) have greatly reduced the weight of power transformers, making the system in Figure 1(c) more attractive from a weight standpoint*.

Since the systems in Figure 1(b) and 1(c) appear to be weight competitive with that in 1(a), there is a strong interest in their application due to their relative simplicity. Both of these systems depend upon a phase controlled voltage regulator for performing all voltage regulation, filtering, and current limiting functions. Thus, it becomes necessary to investigate the characteristics of these functions relative to the requirements of various airborne missions.

The type of voltage regulation circuits that are required appear to be very well developed and have been in use for quite some time (see references 1-4). Not so obvious however, are the implementation of certain current limiting functions and the design of lightweight filters. In the course of this program, a new current limiting circuit was developed and tested for a single phase system. This circuit is feasible for either D.C. or pulsed type loads, and with further development, it can be extended to three phase systems. Filtering requirements have also been investigated, and a computer program for optimizing the weight of LC input filters has been written and tested. Since the ultimate filter design will always be determined by the specific requirements of the source and the load, this area will require further development as these specifications are established.

* Note: If 1(a) and 1(c) both use freon cooled transformers, the weight of the extra components in 1(a) will tend to compensate for the extra transformer weight in 1(c). This compensation effect is usually negligible if conventional transformers are used.


A.C. Generator
Rectifier Filter
High Frequency Inverter Switch
Transformer (Conventional or Freon Cooled)
Rectifier Filter
D.C. Load
low freq. (≈ 400 Hz.)
high freq. (≈ 5-10 KHz.)
Super-conducting A.C. Generator (High Voltage)
Rectifier Filter (Phase Controlled Regulator)
D.C. Load
low freq. (≈400Hz.)
A.C. Generator
Freon Cooled Transformer
Rectifier Filter (Phase Controlled Regulator)
D.C. Load
low freq. (≈400 Hz.)

(a). Conventional method using high frequency inverter.

(b). Simplified system using high voltage superconducting generator.

(c.) Simplified system using low frequency freon cooled transformer.

Figure 1. A.C. to D.C. conversion methods.

II. OBJECTIVES AND SCOPE

The purpose of this research was to find a means of protecting phase control voltage regulators from short circuits while simultaneously minimizing the weight of the LC output filter. In this regard, it was deemed necessary to demonstrate the feasibility of any electronic circuits involved and to develop a systematic method for minimizing filter weight. This information is intended for eventual use as a design tool and as an aid in specifying various performance goals.

III. CURRENT LIMITING CIRCUIT

Theory of Operation

This circuit should be capable of protecting both the source and the load from momentary and long term overloads. It also must be compatible with either a D.C. or a pulsed load, and it should be independent of the output filter characteristics. An electronic circuit to meet these requirements was designed, built and tested with the following specifications:

1. Source Voltage: 25 V.A.C., 60 Hz, single phase

2. Load Voltage: 15 V.D.C.

3. Max. Load Current: 3.75 A.D.C.

4. Overload Trigger Time Delay: 15×10^{-6} sec. (approx.)

5. Momentary Turn-Off Period: 16×10^{-3} sec.

6. Number of Momentary Overloads Required to Produce Permanent Turn-Off: 1 to 4 (adjustable)

7. Maximum Period Between Successive Momentary Overloads to Produce Permanent Turn-Off: 32×10^{-3} sec.

The voltage regulator portion of the circuit is shown in Figure 2(a) and the overload circuit is shown in Figure 2(b).

Voltage regulation is explained by referring to Figure 2(a) (also see page 274, reference 1). The output voltage is sensed across R1 and is compared to the reference voltage CR9 by the differential amplifier formed by Q1 and Q2. The error voltage at the collector of Q1 is proportional to the difference between the output and the reference voltages. This error voltage is applied to the emitter of Q3 to control the delay time of the SCR trigger signal which is derived from this unijunction oscillator. This oscillator is synchronized with the 60 Hz source by its supply voltage which decreases to zero on each half cycle of the 60 Hz source (i.e., the trigger delay time is referenced from the start of each half cycle). For example if the output is too low, the collector voltage at Q1 will increase. This will cause C5 to charge faster; Q3 will fire sooner and the trigger delay time will be less. This means that the SCR's will conduct longer, thus increasing the output voltage until an equilibrium point is reached. A typical waveform of the input voltage across CR3 is shown in Figure 3 (top).

C5 is charged from a sinusoidal source, which is an application of the common ramp and pedestal technique. This method of charging provides a linear gain for the control loop, as described in detail on pages 256-260 of reference 1.

Table I. Parts list for phase controlled voltage regulator with current limiting. Schematic is shown in Figures 2(a.) and (b).

All resistors are 1/4 watt, ±10% unless otherwise noted:

CR1 - CR3: IN347
CR4 - CR7: IN2550
CR8: IN2985B
CR9: IN756A
CR10 - CR11: IN4001
CR12 - CR13: IN751, 5.1V.
CR14: IN4001

R1: 5KΩ, 1/2w., pot.
R2: 2.2KΩ , 7w.
R3: 680 Ω
R4: 470K Ω
R5 - R7: 3.3K Ω
R8: 10K Ω
R9: 4.7K Ω
R10: 22K Ω
R11: 15K Ω
R12: 1M Ω , 1/2w., pot.
R13: 1K Ω
R14 - R15: 33 Ω
R16: 680 Ω
R17: 20K Ω
R18: 22K Ω
R19: 10K Ω , trimpot
R20: 200 Ω
R21: 10K Ω , trimpot
R22: 0.04 Ω
R23: 10K Ω
R24: 10K Ω , trimpot
R25: 2K Ω
R26: 630 Ω
R27: 1.1K Ω
R28: 10K Ω
R29: 10K Ω
R30: 2.2K Ω
R31: 10K Ω , trimpot
R32: 15K Ω
R33 - R34: 330K Ω
R35: 3.3K Ω
R36: 10K Ω
R37: 47K Ω
R38 - R39: 20K Ω
R40: 27 Ω

C1: 22,000 μfd, 50 V.D.C.
C3 - C4: 100 μfd, 30 V.D.C.
C5: 0.1 μfd, 30 V.
C6: 0.001 μfd, 10V.
C7: 2 μfd, 10V.D.C.
C8: 0.1 μfd, 10V.
C9: 4 μfd, 10 V.D.C.
C10: 0.1 μfd, 10V.
C11: 0.01 μfd, 10V.
C12: 0.1 μfd, 30V.
C13: 0.1 μfd, 10V.
C15: 0.1 μfd, 10V.

T1: 1:1:1 Pulse Transformer
T2: 120 V.A.C., 1:1, Isolation Transformer
T3: 120 V.A.C. Variac
L1: 15 m.h., 5A.
Q1 - Q2: 2N3391A
Q3: 2N2646
Q4 - Q5: 2N2222
Q6: 2N2646
Q7 - Q10: 2N2222
Q11: 2N2907
Q12: 2N2222
SCR1 - SCR2: Motorola HEPR 1241
A1 - A2: Fairchild 9601
A3: Signetics 7493
A4: National Semiconductor N7408

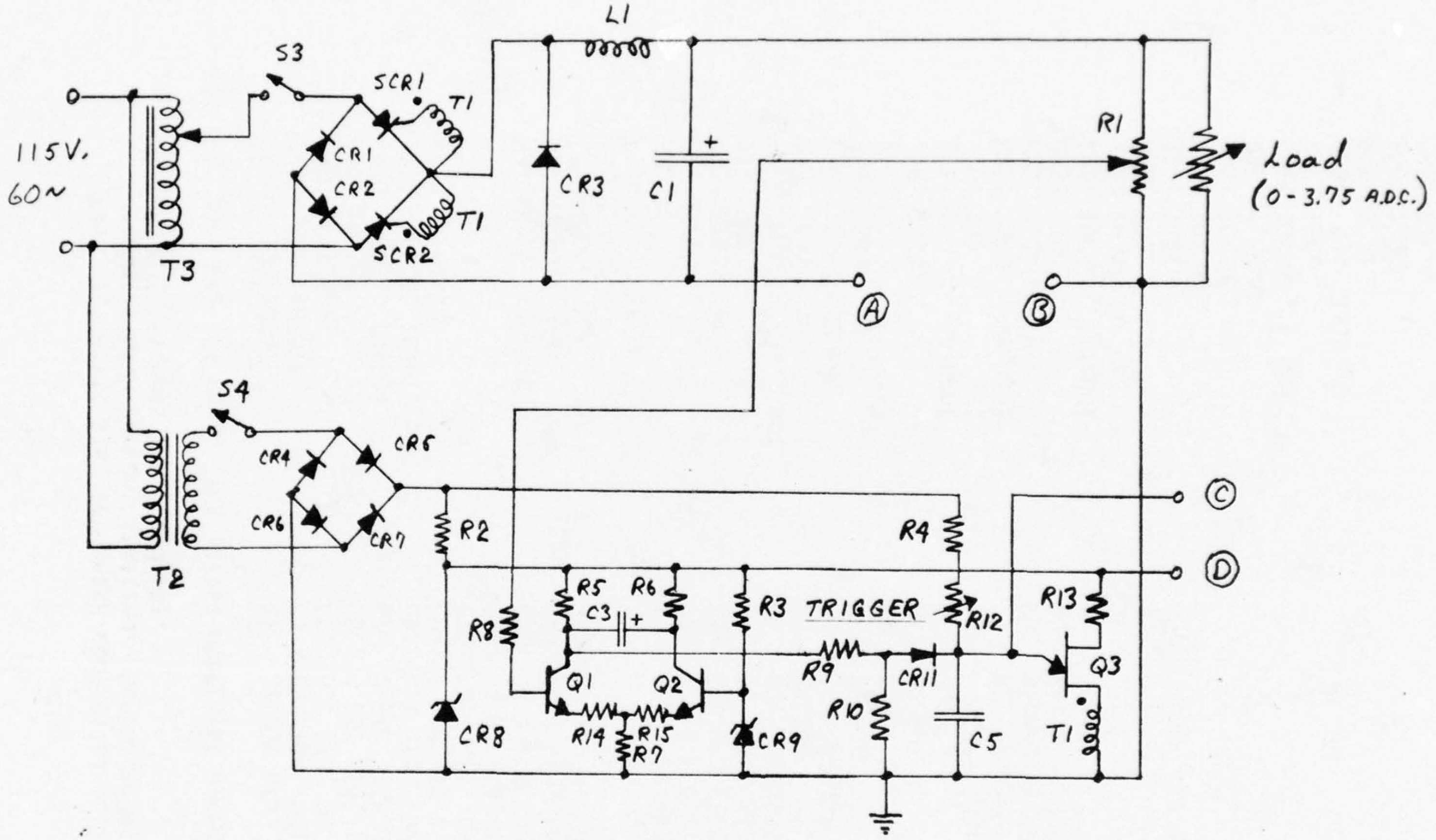


Figure 2(a.). Phase controlled voltage regulator. (See Figure 2(b.) for current overload circuit. See Table I for component values).

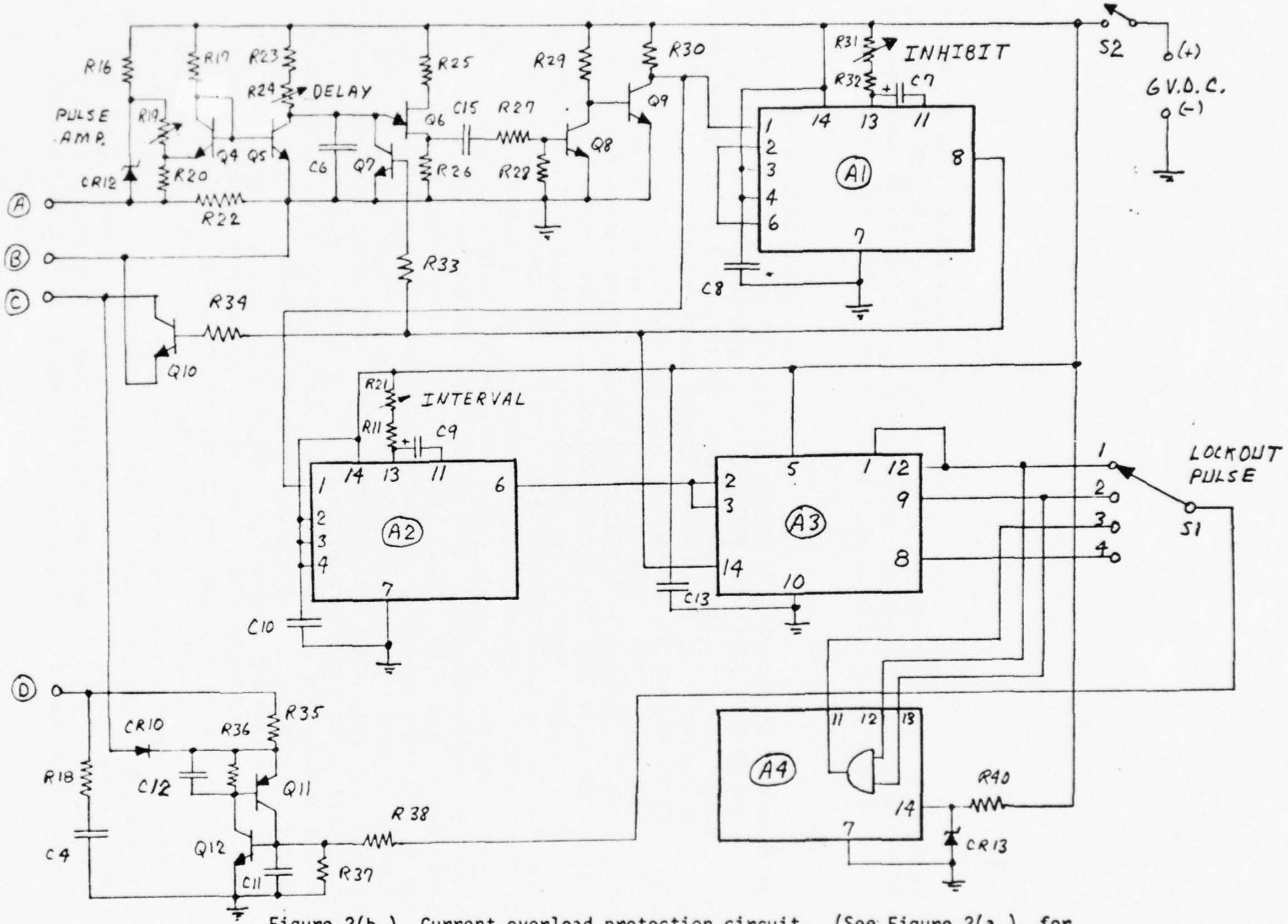


Figure 2(b.) Current overload protection circuit. (See Figure 2(a.). for voltage regulator. See Table I for component values).

Current limiting is performed in the following manner with reference to Figures 2(a) and 2(b). When the voltage across R22 (caused by the load current) exceeds 0.15 V.D.C., Q5 (normally biased into saturation) begins to cutoff, and C6 begins to charge through R23-R24. After a delay of approximately 15 μsec (to allow very narrow current pulses to pass undetected), C6 charges to the required level to fire Q6. The resulting pulse is shaped by Q8 and Q9 and fed to the inputs of A1 and A2. A1 is a simple monostable multi-vibrator that supplies a pulse of 16 m.s. from pin 8 to the base of Q10. This effectively shorts the emitter of Q3 (figure 2(a.)) to ground and prevents the SCR's from firing for at least the next half cycle (two half cycles if the overload occurs just before Q3 would normally fire). A2 is a retriggerable monostable multi-vibrator, meaning that the duration of the output pulse is timed from the occurence of the last input pulse, even if the input pulse occurs when an output pulse is already present. When A2 is triggered (at the same time as A1), it produces an output pulse of 32 m.s. at pin 6. This enables the counter, A3, to count as long as the pulse is present. Simultaneously the output pulse from A1 is applied to the input of A3 at pin 14, and the counter advances by one count. Q7 disables Q6 as long as the A1 output pulse is present. This prevents any further input pulses before A1 has returned to its normal state. After A1 has returned to its normal state, A2 will also revert to its normal state if another input pulse does not occur before the end of the A2 output pulse. When A2 returns to its normal state, A3 resets, and all previous counts are removed. If another input pulse occurs before A2 resets, A1 again disables Q3, and A3 advances by another count. If this completes the maximum allowed number of counts (1 to 4, determined by S1), a pulse appears at pin 11 of A4. This triggers Q11 and Q12, latching these transistors into a permanent conducting state. Since CR10 connects to the emitter of Q3, Q3 is now permanently turned off and can be re-energized only if S4 is opened and closed.

Note: To prevent triggering Q11 and Q12 on initial turn on transients it is necessary to close S2 and S3 before S4. If desired, the requirement for this starting sequence could be eliminated through the use of additional transient suppression circuitry.

Test Results

Once the current limiting circuit in Figure 2 was designed, a breadboard model was built and tested in the laboratory. This circuit met all the requirements specified in the previous section and had the following output voltage regulation characteristics:

Line regulation: +1.7%, 25 VAC to 30 VAC, I_L = 3.5A.

Load regulation: +1.7%, 3.5A to 1.0A, V_{in} = 25 VAC.

No attempt was made to measure the regulation with respect to temperature since a temperature compensated zener was not available for CR9.

Several voltage waveforms of interest were recorded and are shown in Figures 3 through 9. Figures 3 and 4 show some of the steady state waveforms associated with the voltage regulator, while Figures 5 through 9 show some of the pulse waveforms in the current limiting logic circuits. These pulse waveforms are intended to show some of the timing intervals, and they were obtained by applying an external pulse train between the C15-R27 junction and ground. Due to the long time interval involved (up to 112 x 10^{-3} sec. in Figure 9), Figures 8 and 9 had to be taken at fairly low sweep speeds on the oscilloscope. The persistence of the scope screen was fairly low in this sweep range, which resulted in photos of a rather inferior quality. This could have been corrected by using a scope with a high persistence screen, but time did not permit this refinement.

No serious operating problems were noted, although turn on transients require that switches S2 and S3 be turned on before S4 (see Theory of Operation section). In summary, these lab tests indicated that this circuit provides satisfactory current limiting protection, and that the circuit can easily be adjusted to meet specific requirements (current overload levels, timing intervals, etc.). The next logical step in this development would be to design and test a three phase version of this circuit at a higher power level.

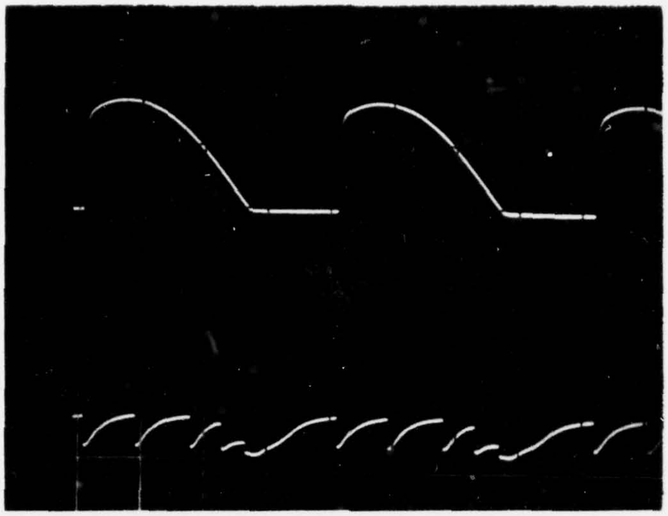


Figure 3 Top: Voltage across CR3, indicating firing delay angle.
Bottom: Voltage across trigger circuit capacitor, C5.
Note waveform during off period of top waveform.
Scale: 2V./cm., 2ms./cm.
Load = 15 V.D.C./3.5A., V_{in} = 25 V.A.C.

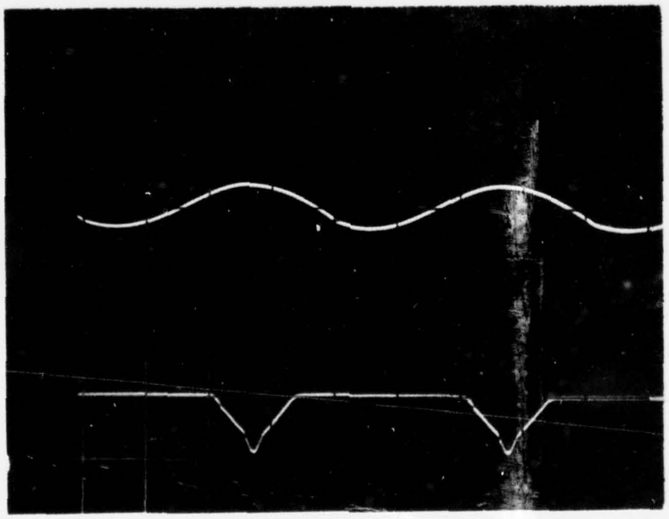


Figure 4 Top: AC ripple voltage across the load.
Bottom: Oscillator supply voltage across CR8.
Note syncrhonization with 60 H_Z input voltage.
Scale: 0.2V./cm. top, 20V./cm. bottom, 2ms./cm.
Load = 15 V.D.C./3.5A, V_{in} = 25 V.A.C.

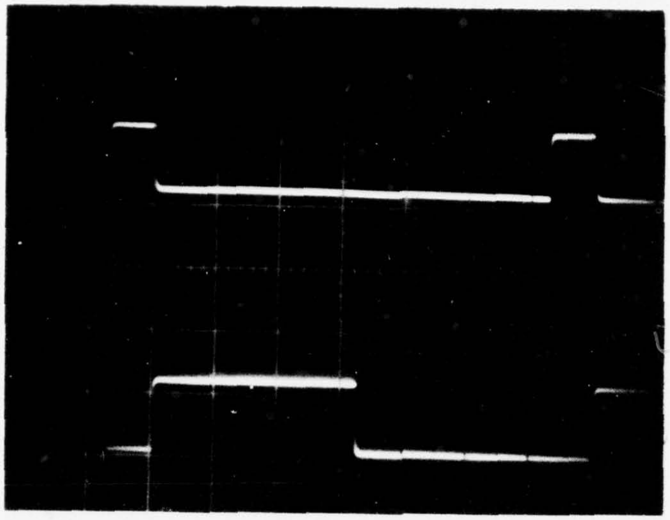


Figure 5 Top: Interval pulse (low value) at A2-6
Bottom: Inhibit pulse (high value) at A1-8
Scale: 5V./cm., 5 ms./cm.

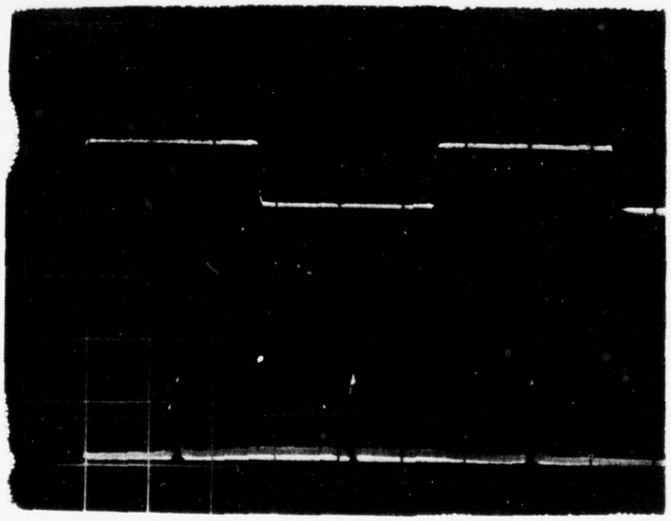


Figure 6 Top: Lockout pulse (high value) at S1-R38 junction.
Bottom: Input pulses at C15-R27 junction. S1 = 1;
note that one pulse occurs before lockout goes positive.
Scale: 5V./cm., 10 ms./cm.

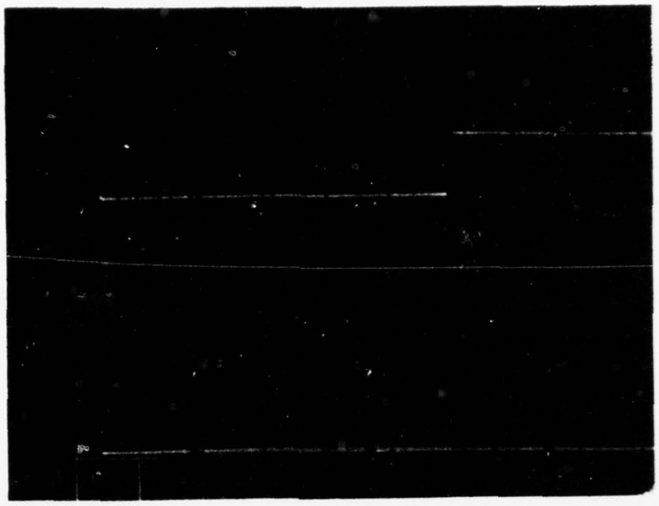


Figure 7 Top: Lockout pulse (high value) at S1-R38 junction. Bottom: Input pulse at C15-R27 junction. S1 = 2; note that two pulses occur before lockout goes positive. Scale: 5V./cm., 10 ms./cm.

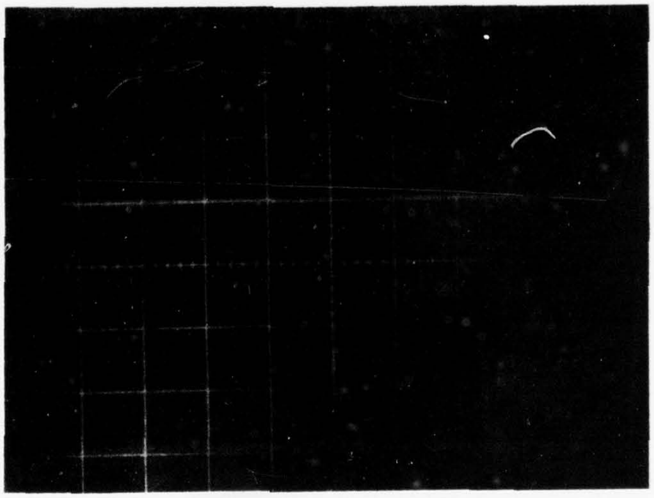


Figure 8 Top: Lockout pulse (high value) at S1-R38 junction. Bottom: Input pulses at C15-R27 junction. S1 = 3; note that three pulses occur before lockout goes positive. Scale: 5V./cm., 10 ms./cm.

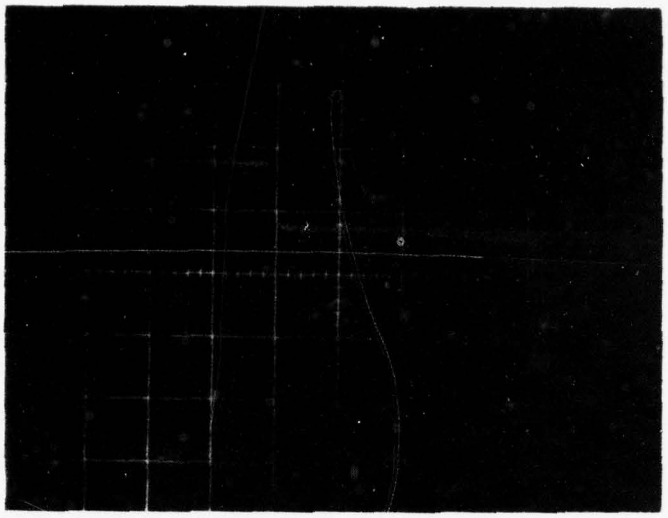


Figure 9. Top: Lockout pulse (high value)at S1-R38 junction. Bottom: Input pulses at C15-R27 junction. S1 = 4; note that four pulses occur before lockout goes positive. Scale: 5V./cm., 20 ms./cm.

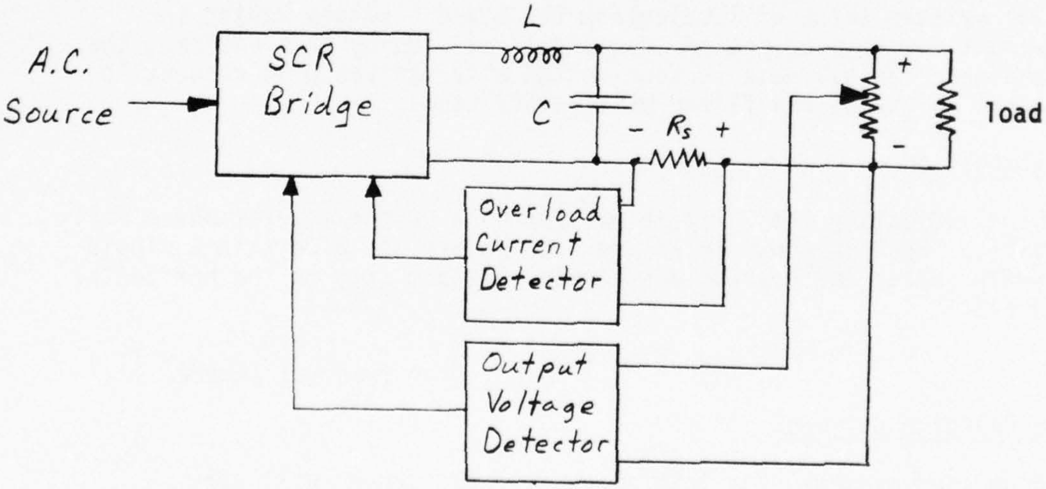


Figure 10. Block diagram of phase control regulator.

IV. LC FILTER DESIGN

Figure 10 shows a simplified diagram of a typical phase control regulator, indicating the SCR bridge, LC filter, output voltage detector and current overload detector. Although simple in form, the LC filter performs several different functions in this system, all of which are quite important. Specifically, this filter must do the following:

1. Attenuate the input ripple voltage to an acceptable level at the load.

2. L must be sufficiently large to limit the capacitor charging current through the SCR's to an acceptable level when the system is first energized.

3. L also must be large enough to limit the current through the SCR's for 1/2 cycle in the event of a short circuit across the load. This will allow adequate time for the current overload circuit to de-energize the bridge before any of the SCR's are damaged.

4. C must be large enough to meet any step load response or maximum output impedance requirements.

Coupled with this is the requirement that the filter must be designed for minimum weight. Due to the conflicting nature of these specifications, the final design will involve certain engineering trade-offs and may require some compromise. To analyze this problem in a systematic fashion, a computer program was written which will calculate the L and C values having the minimum weight for a given set of specifications. Using this program, the systems engineer can evaluate various design alternatives with respect to how the total weight of the filter will be affected.

Ripple Attenuation:

A chart indicating the % ripple vs. $(6\pi f)^2$ x LC for a three phase full wave rectified input is shown in Figure 11. To produce a certain % ripple value, simply select the appropriate ripple constant (k_1) on the horizontal axis, and set

$$(6\pi f)^2 LC \geq k_1 \quad \text{(f = freq. of source)} \quad (1.)$$

Capacitor Charging Current:

With no load present, the initial current at turn-on will vary according to the following expression,

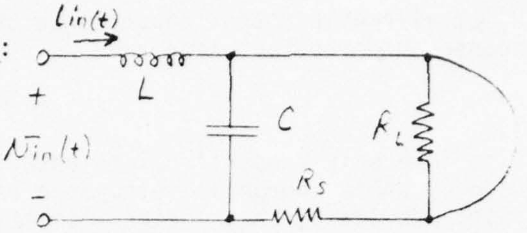
$$i_{in}(t) = V_{in} \times \sqrt{\frac{C}{L}} \sin (Wct) \quad \text{amps} \quad (2.)$$

$$\text{where } Wc = \frac{1}{\sqrt{LC}} \quad o \leq t \leq \frac{\Pi}{Wc} \quad (3.)$$

$$\therefore \quad V_{in} \sqrt{\frac{C}{L}} \leq I_{peak} \quad \text{amps} \quad (4.)$$

where I_{peak} = max. allowable peak input current.

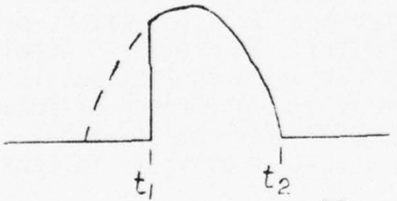
Short Circuit Current:



Ignoring, R_s, $i_{in}(t_2)$ during a shorted condition across the load is given by the following expression:

$$i_{in}(t_2) = \frac{1}{L} \times \int_{t_1}^{t_2} \left[V_{in}(t) \right] dt + i_{in}(t_1) \qquad \text{amps} \quad (5.)$$

Note that $V_{in}(t)$ is not constant (as is approximately true in the usual case) because the current overload circuit prevents the next SCR from firing after the short occurs. Therefore, for a given SCR, the worst case input voltage will have the following form,



$$t_1 = \frac{1}{6f}$$

$$t_2 = \frac{1}{2f}$$

$$i_{in}(t_2) = I_o + \frac{1}{L} \left[\int_{\frac{1}{6f}}^{\frac{1}{2f}} E_{pk} \sin (2\pi\ ft) dt \right] \qquad \text{amps} \quad (6.)$$

where $E_{pk} \approx V_{in}$, I_o = initial current at t_1

$$i_{in}(t_2) = I_o + \frac{V_{in}}{Lf} (0.239)$$

since $i_{in}(t_2) \leq I_{peak}$

$$L \geq \frac{0.239\ V_{in}}{f(I_{peak} - I_o)} \qquad \text{henries} \quad (7.)$$

Step Load Response:

Step load response and maximum power supply output impedance requirements are usually met by limiting the minimum size of the output capacitor. Therefore, these specifications usually can be stated in terms of the following equivalent expression.

$$C \geq C_{min} \qquad \text{farads} \qquad (8.)$$

where C_{min} is the minimum allowable output capacitance that will produce the desired step load response and/or output impedance.

Weight:

Capacitor weight is inversely proportional to both capacitance value and voltage rating. Therefore, where charge and discharge rates are not a factor (as in the case of a constant load), these weight limitations are usually expressed in terms of stored energy per unit weight.

$$\text{Capacitor weight} \approx \frac{CV_{in}^2}{2Dc} \qquad \text{lbs.} \qquad (9.)$$

where D_c = capacitor energy density (joules/lb).

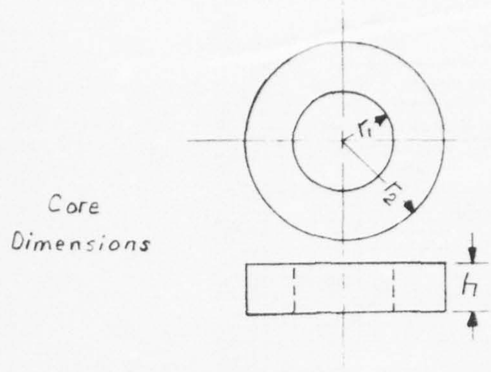
Inductor weight is determined by the amount of wire, core material, and any mechanical encosures or supports. At the D.C. current levels considered in this study (400 to 850 A.D.C.) there do not appear to be any ferrous core materials which will produce a net weight competitive with an air core inductor. This is because the ferrous core would have to be quite large in physical size to prevent total saturation*. Therefore, inductor weight calculations are based exclusively on air core inductors. Time did not permit an extensive investigation into the weight of mechanical supports, but it was presumed that this would not vary significantly over the range of inductance values considered for a given design. Therefore for optimization purposes the weight for a toroidal inductor was expressed strictly in terms of the total conductor weight,

$$\text{Inductor weight} = 2 \times 1.5 \times N \times (r_2 - r_1 + h) A_W D_W \qquad \text{lbs.} \qquad (10.)$$

where N = number of turns
A_W = cross-sectional area of conductor (cm^2.)
D_W = density of copper (0.0196 lbs/cm^3.)
1.5 = winding length factor

The dimensions are indicated in the following diagram:

* Two core manufacturers, Arnold Engineering Company and Magnetics Division of Spang Industries, Inc., were contacted to explore the possibility of new core materials, but these discussions merely supported the above conclusion.



Theoretically, C, N, r_1, r_2 and h can all be varied to produce an optimum total weight. This ignores any mechanical restrictions on the construction of the inductor however, and leads to an expression for the weight which cannot be optimized by analytical methods. An attempt was made to optimize the weight expression using the Newton-Raphson method to solve a set of nonlinear equations. This did not produce any useful results however, since the program tended to converge to an extremum which did not correspond to the minimum weight. Instead of pursuing this course, some additional constraints were imposed to produce a solution which could be found analytically. These constraints were chosen to produce a convenient mechanical shape, and their use should produce a result that is reasonably close to the absolute optimum. The constraints chosen were,

$$h = r_2 - r_1 \qquad \text{cm.} \quad (11.)$$

$$r_2 = 2r_1 \qquad \text{cm.} \quad (12.)$$

$$\therefore \quad \text{Total weight} = \frac{CV^2}{2DC} + 6Nr_1\, A_W\, D_W \qquad \text{lbs.} \quad (13.)$$

Inductor Design:

The inductance of the toroidal air core inductor is given by the following expression,

$$L = (0.133 \times 10^{-8})\, N^2\, r_1 \qquad \text{henries} \quad (14.)$$

Some constraint should also be placed on the minimum size of the core window so that the inductor will be simple to construct. This is usually expressed in terms of the winding factor, WF, which is defined by,

$$WF = \frac{NA_{WC}}{r_1{}^2}\,(1.613 \times 10^{-6}) \qquad (15.)$$

where A_{WC} = conductor cross-sectional area (cir. mils).

Optimizing the Weight of L and C:

The previous design equations are summarized as follows:

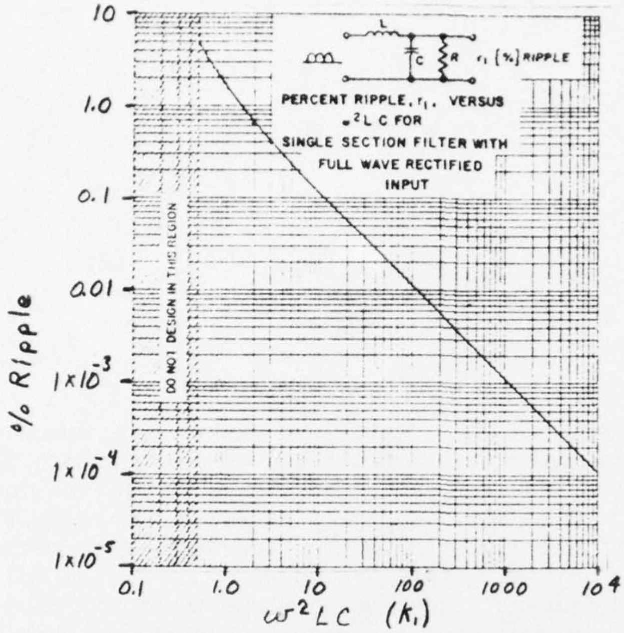


Figure 11. % Ripple vs w^2LC ($w = 6\Pi f$).

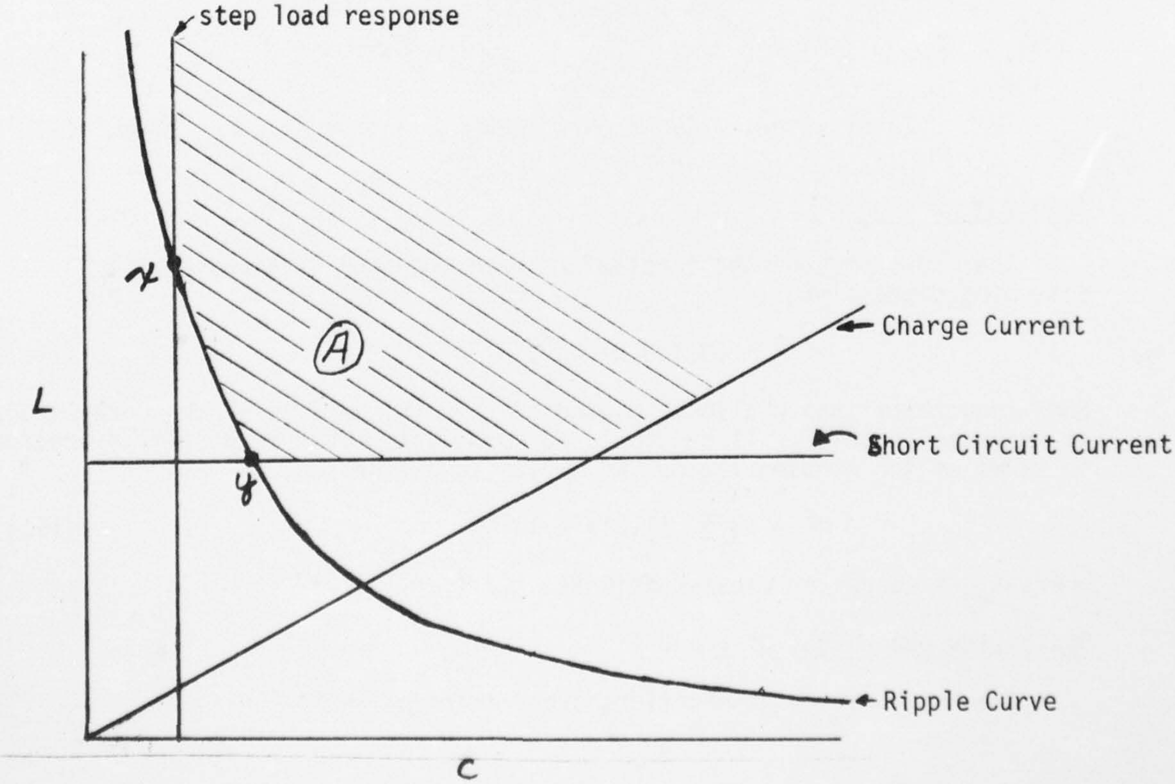


Figure 12. Plot of acceptable L and C values.

$$\text{Total weight} = \frac{CV^2}{2Dc} + 6Nr_1 A_w D_w \qquad \text{lbs.} \quad (16.)$$

$$L = (0.133 \times 10^{-8}) N^2 r_1 \qquad \text{henries} \quad (17.)$$

$$r_1^{\,2} = \frac{NA_{wc}}{WF} (1.613 \times 10^{-6}) \qquad \text{cm}^2 \quad (18.)$$

$$L \geq \frac{k_1}{(355.31) f^2 C} \qquad \text{henries} \quad (19.)$$

$$L \geq \left(\frac{V_{in}}{I_{peak}}\right)^2 C \qquad \text{henries} \quad (20.)$$

$$L \geq \frac{0.239 V_{in}}{f (I_{peak} - I_o)} \qquad \text{henries} \quad (21.)$$

$$C \geq C_{min} \qquad \text{farads} \quad (22.)$$

Figure 12 shows an example plot* of L versus C for each of the four constraint equations, 19-22. The shaded area, A, indicates the range of L and C values that will satisfy all of the constraints. Since L and C are inversely proportional only in equation (19.), this is the only constraint equation that indicates where a compromise can be made between L and C. Therefore, the following method was used to find the minimum weight:

1. Find the L and C that produce the minimum value of (16.) using (17.) - (19.) as equality constraints.

2. Check to determine if the inequality constraints (20.) - (22.) are satisfied.

3. If necessary, alter L and/or C until (20.) - (22.) are satisfied. For example, suppose L must be increased to satisfy (21.). It may then be possible to decrease C according to (19.), provided that C always satisfies (20.) and (22.). For the particular example shown in Figure (12.), the range of acceptable values lie between points x and y on the ripple curve.

* It should be stressed that Figure 12 is only an example, and that the intersection points of the curves may vary considerably for different examples.

The optimization of (16.) involves four variables, C, L, N, and r_1 and three equality equations (17.), (18.), and (19.). Since the variables are restricted to positive values, it is possible to solve for C, L, and N in terms of r_1, and then substitute these values in (16.). The results of this operation are as follows:

$$\text{let } L = A_1 N^2 r_1, \; LC = A_2, \; r_1^2 = A_3 N, \qquad \text{Weight} = A_4 C + A_5 N r_1 \qquad (23.)$$

Where A_1 through A_5 can be determined by referring to equations (16.)-(19.).

$$\therefore \text{ Weight} = \frac{A_6}{r_1^5} + A_7 r_1^3 \quad \text{lbs.} \qquad (24.)$$

$$\text{where } A_6 = \frac{A_4 A_2 A_3^2}{A_1}, \; A_7 = \frac{A_5}{A_3}$$

$$\frac{d(\text{weight})}{dr_1} = \frac{-5A_6}{r_1^6} + 3A_7 r_1^2 = 0 \qquad (25.)$$

$$\therefore \; r_{1(\text{opt.})} = \left(\frac{5A_6}{3A_7}\right)^{1/8} \quad \text{cm.} \qquad (26.)$$

The corresponding values of weight, L, C, and N now can be found by making the appropriate substitutions.

Computer Program for L-C Weight Optimization:

A listing of the L-C weight optimization program, OPFIL, N01 (Optimum Filter) is given in Appendix I. This program is designed for use in an interactive mode, and will analyze as many different case studies as desired in a single run.

An example study is shown on the following pages along with the comments listed below:

1. User types 1 to begin the next study or 2 to end the run.

2. The user types in the indicated information upon request from the program.

3. Minimum inductance for short circuit protection.

4. Optimum L and C data, ignoring surge current and step load response constraints (these constraints are checked later in the program).

5. Peak short circuit and charging currents, using the final values of L and C.

6. Number of strands of #8 wire required for the conductor. Only #8 wire is used in the design.

7. User types 1 to begin the next study or 2 to end the run.

It should be noted that some case studies will print out additional information, depending on how often the L and C values have to be changed to satisfy all of the constraints. This additional data will always be accompanied with an explanation, and the final values for L and C will always appear last.

```
    RUN,FTN

             .816 CP SECONDS COMPILATION TIME
①   TYPE 1 TO CONTINUE, 2 TO END1

    POWER=(MW),VOLTAGE=(KV),FREQ=(1 PHASE,HZ)
    CAP ENERGY DEN=(J/LB),CUR DEN=(CIR MIL/AMP)
②   RIPPLE CONSTANT
    SURGE CUR=(AMP),WINDING FACT=(%)
      MIN CAP=(UFD)25,60,400,50,400,10,10000,60,0

③    L MIN FOR SHORT=   .374E+01MH
④    L=   .833E+01MH C=   .196E+02UFD N= 624TURNS
     CWT=   .704E+03LBS LWT=   .117E+04LBS R1=   .684E+01IN R2=   .137E+02IN
⑤    SHORT I=   .440E+04AMPS CHARGE I=   .280E+04AMPS
     NO. STRANDS #8 WIRE=  10
⑥    CURRENT DENSITY=   .432E+03CIR MILS/AMP
     WIRE LENGTH=   .256E+05INCHES
     RESISTANCE=   .134E+00OHMS
⑦   TYPE 1 TO CONTINUE, 2 TO END
```

V. CONCLUSIONS AND RECOMMENDATIONS

In summary, the primary goals accomplished in this study were:

1. To design and demonstrate the feasibility of a current overload protection circuit that can possibly be extended to high power (25 to 50 MW) phase control voltage regulators.

2. To develop a systematic method for calculating and minimizing the weight of high power LC filters, subject to various design constraints.

One point that continues to arise in this work is that the technology of lightweight power conditioning systems has tended tr lag that of lightweight power sources, such as superconducting generators. As a result, systems engineers are faced with the propsect of a power conditioning system which may weigh at least three times the combined weight of the turbine-generator combination*. Not only do these systems have a relatively high weight, but they also are difficult to protect in the event of a short circuit across the output. The point to be made, is that certain design problems such as those discussed here will need further attention before power conditioning systems achieve a level of performance that is consistent with the rest of the power system.

In that light, this study should be regarded as preliminary in scope, and further efforts should be made to extend and apply these results to actual high power systems. Specifically, the single phase current limiting circuits discussed here should be extended to three phase systems, and the LC filter computer program should be utilized to collect more data on the trade-offs between weight, reliability, and electrical characteristics. If the current overload circuit is properly coordinated with the LC filter design, it should be possible to further minimize the filter weight without an undo sacrifice in electrical performance.

* The turbine for a 25 MW system is expected to weigh approximately 300 lbs., while reference 6 estimates a 25 MW superconducting generator to weigh 2160 lbs. State-of-the-art power conditioning systems weigh around 4 lbs/KW, while those using the most recent technology may reach 0.3 lbs/KW. Thus, even the best projected 25 MW power conditioning system will weigh 7500 lbs. as compared to a turbine generator set which will weigh only 2460 lbs.

VI. REFERENCES

1. SCR Manual, Fifth Edition,Semiconductor Products Department, General Electric Company, Syracuse, NY, 1972.

2. SCR Designers Handbook, First Edition, Semiconductor Division, Westinthouse Electric Corporation, Youngwood PA, 1963.

3. Semiconductor Controlled Rectifiers - Principles and Applications of p-n-p-n Devices, F.E. Gentry, et al, Prentice Hall, Incl, Englewood Cliffs, NJ, 1964.

4. Thyristor Phase - Controlled Converters and Cycloconverters, B.R. Pelly, Wiley, New York, NY, 1971.

5. J.L. McCabria and C.C. Kouba, "Features of a High Voltage Airborne Superconducting Generator," Proceedings of the National Aerospace Electronics Conference 1973, Institute of Electrical and Electronics Engineers, Inc. Publication 73 CH0735-1 AED, pp. 216-219, Dayton OH, May 1973.

6. J.L. McCabria, R.D. Blaugher, and J.H. Parker, Jr., "Superconducting Generator Development," Proceedings of the IEEE 1975 National Aerospace and Electronics Conference, Publication 75 CH0956-3 NAECON, pp. 261-271, Dayton OH, June 1975.

7. J.P. Welsh, D.L. Lockwood, R.L. Haumesser, R.I. McNall, Jr., and D.L. Pierce, "The Development of Lightweight Transformers for Airborne High Power Supplies - A Status Report," Proceedings of the IEEE 1975 National Aerospace and Electronics Conference, Publication 75 CH0956-3 NAECON, pp. 272-279, Dayton, OH, June 1975.

VII. ACKNOWLEDGEMENT

The author would like to acknowledge the efforts of several personnel at AFAPL/POD who were quite helpful during the course of this research. Of particular importance were the advice and comments of P. C. Herren, P. E. Stover and R. L. Verga.

APPENDIX I. LC Filter Weight Optimization Program

```
       PROGRAM OPFIL(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
       REAL JLB,IO,LMIN,L,LWT,L1,L2
1      WRITE(6,2)
2      FORMAT(*0*,*TYPE 1 TO CONTINUE, 2 TO END*)
       READ(5,*) IV
       IF(IV.GE.2) GO TO 3
       WRITE(6,4)
4      FORMAT(*0*,*POWER=(MW),VOLTAGE=(KV),FREQ=(1 PHASE,HZ)*)
       WRITE(6,49)
49     FORMAT(*0*,*CAP ENERGY DEN=(J/LB),CUR DEN=(CIR MIL/AMP)*)
       WRITE(6,48)
48     FORMAT(*0*,*RIPPLE CONSTANT*)
       WRITE(6,44)
44     FORMAT(*0*,*SURGE CUR=(AMP),WINDING FACT=(%)*)
       WRITE(6,47)
47     FORMAT(*0*,2X,*MIN CAP=(UFD)*)
       READ(5,*) P,V,F,JLB,CMA,RIP,SI,WF,CMIN
C FIND WIRE SIZE AND AREA
       P=P*1000000
       V=V*1000
       IO=P/V
       WF=WF/100
       CMIN=CMIN/1000000
       CM=CMA*IO
       AM=(CM/18000)
       AAM=AM*2
       M=AAM
       M=M/2
       BM=M
       IF(AM.GT.BM) M=M+1
       AWCM=M*18000
       AWCC=AWCM*(0.51 E-05)
C  FIND MIN L FOR SURGE
       LMIN=(0.239)*V/(F*(SI-IO))
       L2=LMIN*1000
       WRITE(6,18) L2
18     FORMAT(*0*,1X,*L MIN FOR SHORT=*,E10.3,*MH*)
C  FIND OPT LC WT FOR R2=2*R1
       A1=0.133E-08
       A2=RIP/(355.31*F*F)
       A3=AWCM*(1.61E-06)/WF
       A4=V*V/(2*JLB)
       A5=6*AWCC*(0.0196)
       A6=A4*A2*A3*A3/A1
       A7=A5/A3
       R1=(5*A6/(3*A7))**0.125
       R2=2*R1
       N=R1*R1/A3
       L=A1*N*N*R1
       C=A2/L
```

```
      CWT=A4*C
      LWT=A5*N*R1
      R1=R1/2.54
      R2=R2/2.54
      L1=L*1000
      C1=C*1000000
      WRITE(6,12) L1,C1,N
12    FORMAT(*0*,1X,*L=*,E10.3,*MH*,1X,*C=*,E10.3,*UFD*,1X,*N=*,
     1 I4,*TURNS*)
      WRITE(6,13) CWT,LWT,R1,R2
13    FORMAT(*0*,1X,*CWT=*,E10.3,*LBS*,1X,*LWT=*,E10.3,*LBS*,
     1 1X,*R1=*,E10.3,*IN*,1X,*R2=*,E10.3,*IN*)
C  CHECK CHARGING CURRENT
      TEST=V*((C/L)**0.5)
      IF(TEST.LE.SI) GO TO 10
      WRITE(6,8)
8     FORMAT(*0*,*CHARGE CURRENT TOO HIGH, USE L/C CONSTRAINT*)
      L=V*((RIP**0.5)*(0.053)/(SI*F))
      C=((SI/V)**2)*L
17    CONTINUE
      R1=(L*A3*A3/A1)**0.2
      R2=2*R1
      N=R1*R1/A3
      CWT=C*V*V/(2*JLB)
      LWT=6*N*(R2-R1)*AWCC*(0.0196)
      L1=L*1000
      C1=C*1000000
      R1=R1/2.54
      R2=R2/2.54
      WRITE(6,12) L1,C1,N
      WRITE(6,13) CWT,LWT,R1,R2
10    CONTINUE
C  CHECK SHORT CURRENT
      IF(L.GE.LMIN) GO TO 15
      WRITE(6,16)
16    FORMAT(*0*,*SHORT CKT CURRENT TOO HIGH, USE LMIN*)
      L=LMIN
      C=RIP/((355.31)*F*F*L)
      GO TO 17
15    CONTINUE
C  CHECK FOR MIN CAP
      IF(C.GE.CMIN) GO TO 30
      WRITE(6,38)
      C=CMIN
38    FORMAT(*0*,1X,*C TOO LOW,  USE CMIN*)
      L=A2/C
C  CHECK CHARGING CURRENT
      TEST=V*((C/L)**0.5)
      IF(TEST.LE.SI) GO TO 37
      WRITE(6,8)
      L=C*((V/SI)**2)
37    CONTINUE
      R1=(L*A3*A3/A1)**0.2
      R2=2*R1
      N=R1*R1/A3
      CWT=C*V*V/(2*JLB)
      LWT=6*N*(R2-R1)*AWCC*(0.0196)
      L1=L*1000
      C1=C*1000000
      R1=R1/2.54
      R2=R2/2.54
      WRITE(6,12) L1,C1,N
      WRITE(6,13) CWT,LWT,R1,R2
31    CONTINUE
C  CHECK SHORT CURRENT
      IF(L.GE.LMIN) GO TO 30
      WRITE(6,16)
      L=LMIN
      GO TO 37
30    CONTINUE
      SI=IC+(0.239)*V/(F*L)
      CI=V*((C/L)**0.5)
      WRITE(6,19) SI,CI
19    FORMAT(*0*,1X,*SHORT I=*,E10.3,*AMPS*,1X,*CHARGE I=*,
     1 E10.3,*AMPS*)
      WRITE(6,20) M
20    FORMAT(*0*,1X,*NO. STRANDS #8 WIRE=*,I4)
      CD=(M*18000)/IO
      WRITE(6,21) CD
21    FORMAT(*0*,1X,*CURRENT DENSITY=*,E10.3,*CIR MILS/AMP*)
      WL=N*6*(R2-R1)
      WRITE(6,22) WL
22    FORMAT(*0*,1X,*WIRE LENGTH=*,E10.3,*INCHES*)
      OHMS=(WL/12)*(0.00063)/M
      WRITE(6,23) OHMS
23    FORMAT(*0*,1X,*RESISTANCE=*E10.3,*OHMS*)
      GO TO 1
3     STOP
      END
```

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT-PATTERSON AFB, OHIO

&

EGLIN AFB, FLORIDA

(Conducted by Auburn University)

A STUDY ON NUMERICAL METHODS FOR COMPUTING TRANSONIC FLOWS IN TURBOMACHINES

Prepared by:	Shu-Yi S. Wang, Ph.D
Academic Rank:	Associate Professor
Department & University:	Mechanical Engineering University of Mississippi
Assignment:	AF Aero Propulsion Laboratory Turbine Engine Division Components Branch
USAF Research Colleague:	Kervyn D. Mach, Ph.D
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

A STUDY ON NUMERICAL METHODS FOR COMPUTING TRANSONIC FLOWS IN TURBOMACHINES

by

SHU-YI S. WANG

ABSTRACT

The objectives of this study are to improve the efficiency of the available computer program for simulating the transonic flow of a viscous, compressible fluid through a cascade of blades in axial flow turbomachines, and to examine the various, newly developed numerical technique in order to identify the one(s) with promising potential to handle more realistic turbomachine blade element flows in the future.

The improvement of an existing computer program for transonic cascade flow analysis has been carried out in two steps: the first step was to simplify the mathematical model or governing differential equations, and the second step was to develop a better solution technique. By order of magnitude analysis, the original differential equations were simplified and a reduction of 30% in computing time was achieved in some cases, however, it was not consistent for all cases.

The ADI (Alternating-Direction Implicit) difference method was chosen as the solution technique to replace the explicit MacCormack finite difference scheme used in the original program. The formulation and programming of the numerical solution scheme have been completed. From the experience of trial runs, it is estimated that the computation efficiency of the new program (ADI) may not be much better than the explicit version. Furthermore, the idealizing and simplifying assumptions of the mathematical model used in the original program have limited the extension or generalization of the model to cover more realistic flows. Therefore, the development of new solution approaches is needed.

The finite element or finite-difference/finite-element technique has been proven to have potential in simulating the transonic flows through a cascade of blades in a turbomachine. Further studies in this area is recommended.

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Dr. Kervyn Mach and Dr. James S. Petty for their continuous interest, valuable suggestions, consultation and assistance during the entire course of the present work.

Appreciation is also extended to Ms. Kathy Nagy for typing the manuscript, and the author's wife, Jine, a Ph.D candidate in Physics, for verifying the formulation of the ADI equations.

The financial support of the AFOSR and the administrative assistance of the Turbine Engine Components Branch of the AF Aero-Propulsion Laboratory are also acknowledged here.

The author also wishes to express his appreciation to Mr. J. Fred O'Brien, the director of this program for organizing this very meaningful program and for his outstanding effort to make the program a success.

AD-A031 017 AUBURN UNIV ALA SCHOOL OF ENGINEERING F/G 5/2

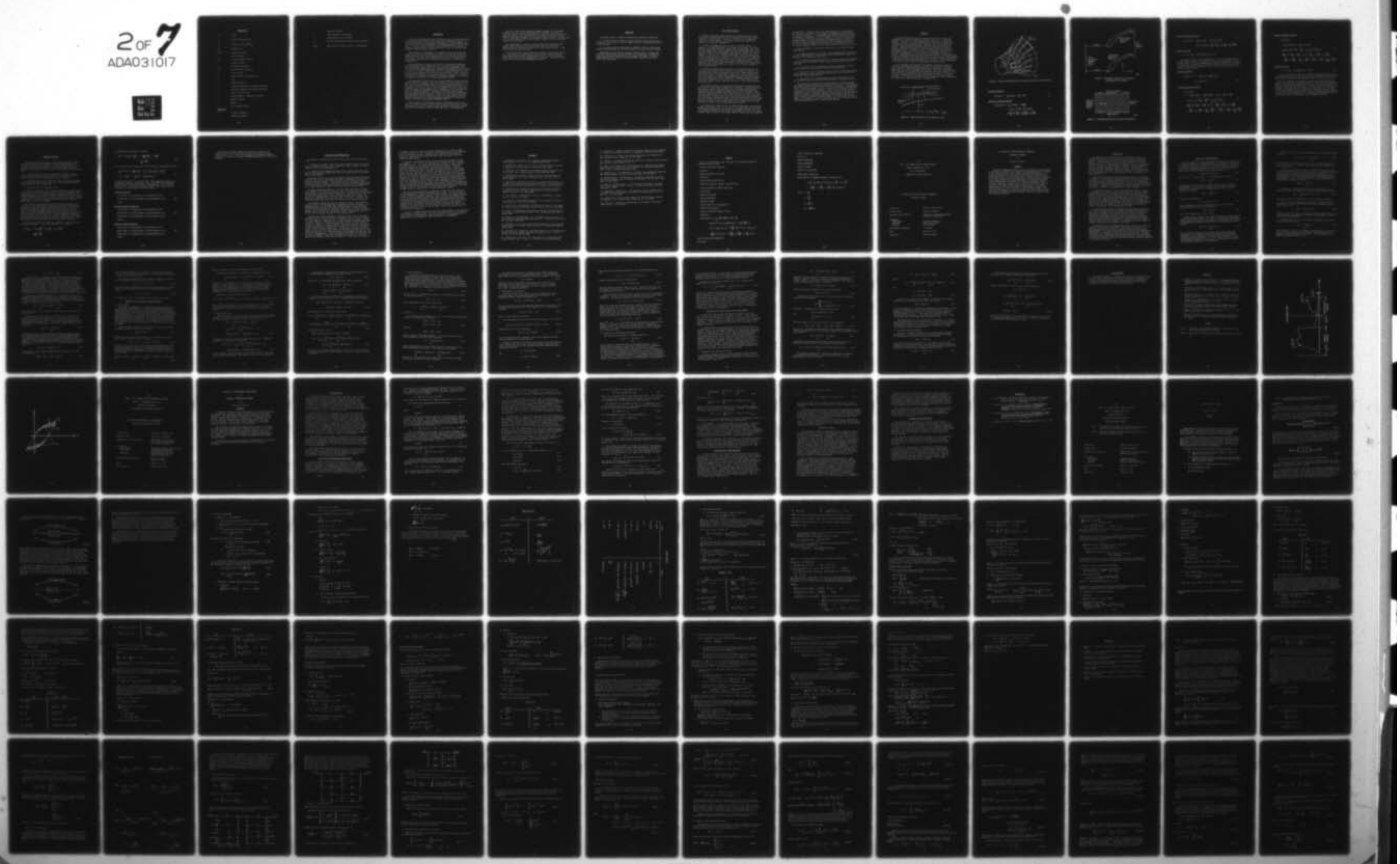
1975 USAF/ASEE SUMMER FACULTY RESEARCH PROGRAM.(U)

SEP 75 J F O'BRIEN F44620-75-C-0031

UNCLASSIFIED AFOSR-TR-76-1113 NL

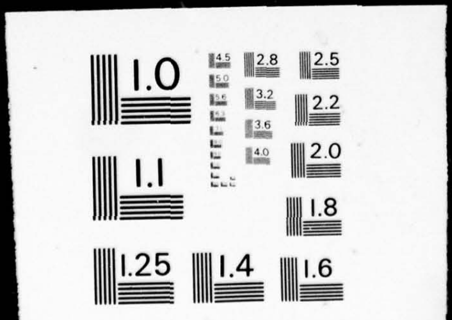
2 OF 7

ADA031017



2 OF 7

ADA031017



NOMENCLATURE

B	$1/\gamma M_r^2$
C_1	$(d\theta_1/dx + y d\theta_{21}/dx)/\theta_{21}$
C_2	$(d^2\theta_1/dx^2 + y d^2\theta_{21}/dx^2)/\theta_{21}$
C_3	$(d\theta_{21}/dx)/\theta_{21}$
$F(m)$	stretching function
h	static enthalpy
H_R	rothalpy $(h+W^2/2-(R\omega)^2/2)$
l	reference length
m	meridional coordinate
N_R	Reynolds number $(\rho_r W_r l/\mu_r)$
p	static pressure
R	radius measured from engine axis
t	time
W	relative velocity
x	meridional coordinate in computational plane
y	tangential coordinate in computational plane
γ	ratio of specific heats
θ	angular measure in tangential direction
μ	dynamic viscosity
ρ	density
ω	rotor angular velocity

SUBSCRIPTS

m	meridional component
θ	tangential component

r	reference condition
1	lower boundary of flow passage
2	upper boundary of flow passage
21	passage spacing between upper and lower boundaries

NOTE: Additional definitions are given in the Appendix

INTRODUCTION

Ever increasing demands of higher efficiency as well as better performance turbine engines have accelerated the research and development efforts in various areas related to the turbine engine design. The fluid dynamic characteristics within the turbomachines, including compressors, turbines, fans, etc., is the most essential aspect of the turbomachine design, because it effects all other aspects of design consideration, such as structure, heat transfer, stability, noise, and so on.

Traditionally, the development of turbomachines has been carried out primarily by empirical techniques with little assistance from theoretical analysis based on either one-dimensional flow or two-dimensional potential flow theory. At times, when the medium was water or air flowing at low speed, those theory generally provided useful guidance to the designers of turbomachines.

As the performance as well as the efficiency requirements become more and more demanding, the turbomachines have to be operated at much higher inlet speed and/or temperature as well as much heavier aerodynamic loads. The real flow phenomenon within the turbomachines has become extremely complex. It is not only viscous, compressible, and truly three-dimensional, but also non-uniform, unsteady, and in most cases non-equilibrium. Besides, the complicated boundary geometry and conditions add more difficulties. Therefore the classical, over-simplified theories have become inadequate for providing the essential information to the modern turbomachine design.

The experimental investigations on the new turbomachine designs are very expensive and time-consuming. The cost of failure of a new turbine design during the engine development phase has reached staggering proportions. Not only will it require millions of dollars to design, manufacture and test a replacement, but the lost time, which often amounts to a year or more, can seriously impede the development cycle. Furthermore, the increasingly wide variety of operating conditions and blade configurations make a purely empirical approach impossible. Consequently, the analytical investigations have become more and more important. They are not only needed to provide some preliminary design information as well as to intelligently guide the hardware experiments, but to perform computer experiments for feasibility studies on a closely approximated mathematical model.

With the advent of the numerical solutions techniques as well as the capacity of digital computers, more and more realistic models may be solved. The present brief study is devoted to the improvement of numerical solution techniques for solving the blade-to-blade flow within the turbomachines.

As a first step the best available computer program for solving blade-to-blade flow of a viscous, compressible fluid developed by the Detroit Diesel Allison Division of General Motors [1], under Air Force Aero-Propulsion Laboratory contract, have been improved by eliminating a few more viscous terms of lower order of magnitude In the Navier-Stokes equation. The computation time has been reduced by 30% in certain cases, however, it is not true consistantly for all cases.

The second step is to apply the method of ADI (Alternating-Direction Implicit) Approximation for solving the same problem to replace the MacCormach finite differencing scheme used in the original program. It may further reduce the computing time.

A review of various existing numerical techniques has also been carried out. The feasibility of application of each of them to the solution of the blade-to-blade flows has been examined carefully. The Finite Element Method seems to be the best candidate for this purpose. The reason will be given in the Conclusion and Recommendation section.

OBJECTIVES

The present study is intended to achieve the following two objectives:

1. To improve the best computer program available to the Aero-Propulsion Laboratory, if possible, so that the computing time is reduced without sacrificing the accuracy.

2. To study various existing numerical methods in order to explore some promising new techniques for solving the same or even more realistic problems.

The problem to be solved is a steady, transonic flow of a viscous, compressible fluid through a cascade of blades of turbomachines. Although the axial flow compressor has been used as an example, the method, or computer program may be easily modified to obtain flowfield properties within the axial turbine and pumps.

THE STATE-OF-THE-ART

The many and varied methods of analytic calculation used to predict the flows in turbomachines are briefly reviewed here due to the limited length of this short report. Attention is centered on techniques for 'analyzing' a given geometry of the blade-to-blade flow rather than methods for 'designing' turbomachines.

It must be re-emphasized that the flow is extremely complex. In general, it is three-dimensional, non-uniform, and unsteady. The fluid should be considered as viscous, compressible and heat-conducting. When the inlet and/or wheel speed is high, the shock wave boundary layer interactions, separation and reattachment, heat transfer effects, etc. are all important. Even with the remarkable advancement of computational fluid dynamics in recent years, the solution of "real" flow within turbomachines is still beyond our reach. However, it is always true that the properly simplified mathematical models usually give us some significant prediction of the flow phenomena at least to the order of magnitude accuracy. And, sometimes, most of these models and solution techniques may be successfully refined to give better accuracy at a higher cost in terms of manhour as well as computing time.

The most successful attack on the flow in turbomachines has been to solve the flow separately in two families of intersection surfaces - the blade-to-blade surfaces through which the aerofoil shaped blades project, and the meridional surfaces formed by taking radial planes through the axis of the machine. Circumferential averaging of the equation of motion across the blade pitch yields an equivalent axisymmetric problem solvable in a single representative meridional plane. Information obtained from solutions in blade-to-blade planes must be used in the meridional solution and vice versa.

Most useful analyses of the blade plane problem have assumed that the flow relative to the blade is steady and irrotational. Classical solutions obtained by conformal transformation, such as Merchant/Collar [2] and Gostelow [3], and by singularity methods, such as Schlicting/Scholz [4], Isay/Marteensen [5], and Wilkinson [6], have been most successful in the study of incompressible flow. The singularity methods, in which the blades are replaced by vortex and soure-sink distributions, have been extended to deal with subsonic compressible flow, firstly by compressibility corrections, such as von Karman [7] and Tsien [8], and secondly by representing the compressibility terms by a source-like function in the equations and solving the flow iteratively, such as Imbach [9] and Price [10].

For compressible steady flow in the blade-to-blade plane, two numerical methods are now widely used. In the first, the streamline curvature method, a differential equation for the gradient of streamwise velocity along the normal (Bindon/Carmichael [11]) or near normal (Katsanis [12]) to the streamline [illegible] written in terms of the assumed radius of curvature of the streamline.

This equation is integrated across the blade passage to give the velocity and hence density profiles. The constant of integration is determined by the continuity equation. The streamlines are redetermined and the solution is repeated to convergence. The method appears to give satisfactory answers for isentropic transonic flow, but its validity in a flow with shocks must be open to doubt. In the second method, the same equation is written in terms of a stream function satisfying the continuity equation, and solved by finite difference schemes of matrix inversion or relaxation [14, 15]. It has been extended to deal with isentropic transonic flow by Perkins [16].

Recent developments in the calculation of flow in the blade-to-blade plane include:

1. A numerical solution to the flow in the hodograph plane, involving an eliptic type solution in the subsonic region and a matched characteristic solution for the supersonic region [17].

2. A streamline curvature calculation for subsonic region and a characteristics solution for supersonic region, the two being matched at the sonic line [18].

3. A solution of the time-dependent hyperbolic equations in order to calculate mixed sonic and supersonic flow with shocks [19-22].

4. The explicit finite difference solution of the time-dependent Navier-Stokes equations in quasi-conservative form governing the transonic flow of viscous and compressible fluid [1].

5. The use of finite element method for steady incompressible flow with possible extension to unsteady flow [23] as well as compressible flow.

The present brief study during the past 9 weeks is trying to improve the computer program developed by Kurzrock and Novick [1] of General Motors under contract with the Air Force Aero-Propulsion Laboratory. The MacCormack finite differencing scheme was used in solving the nonlinear Navier-Stokes equations and the continuity equation. Since it is an explicit scheme, the convergence is slow and the magnitude of time interval is limited by the stability criterion. Therefore, the numerical technique of ADI (Alternating Direction Implicit) Approximation Method has been chosen to replace the explicit MacCormack finite difference scheme. It is hoped that the convergence will be improved in order to save computing time. The formulation and solution procedures are given in the next two sections.

ANALYSIS

As briefly described in previous sections that the flow through the turbomachines, including compressors, turbines, fans, etc., is extremely complex. In general, it is three-dimensional, compressible, viscous and unsteady. Besides, the geometry within which it flows is complicated and the boundary conditions are involved. When the inlet speed reaches to transonic range, the flowfield becomes even more complicated. There are subsonic, transonic and supersonic regions existing within the flowfield and where they are is not known a priori. At the present state of the art the complete solution of the real transonic flow through a cascade of blades of a turbomachine is not in existence as surveyed in the previous section. Among the existing solutions of simplified models, the one developed by Kurzrock and Novick [17] of Detroit Diesel Allison Division of General Motors is the best available one at the present time, because it includes the dominant viscous terms in the Navier-Stokes equations to account for losses generated in the flow field as a result of shock waves and viscous mixing.

In this approach, the nonlinear time-dependent conservation equations (continuity and momentum equations) are used. By assuming the constant total enthalpy, or Rolthalpy as defined by some investigators, the energy equation is not used for mathematical simplicity. This set of equations are first formulated in the Meridional and Tangential Coordinates (Figures 1 & 2) and simplified by assuming that flow field variations normal to the stream surfaces are ignored as well as that the locations of the stream surfaces are known. The resulted governing differential equations, after normalization, are given below. Details are given in reference [1].

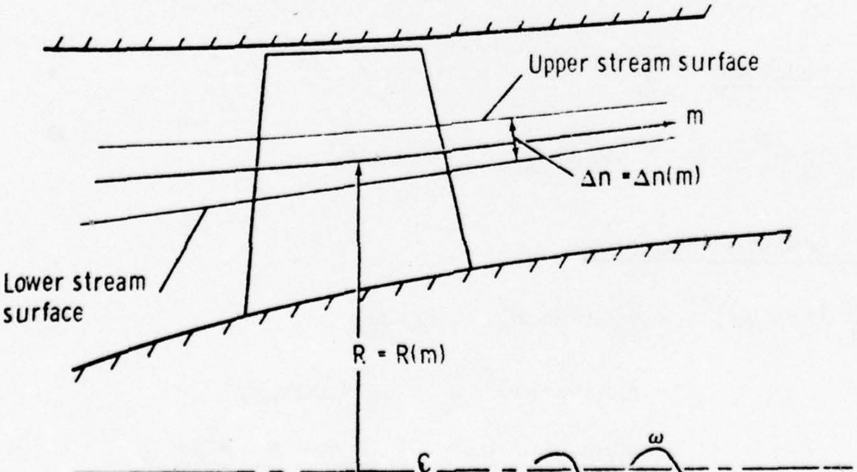


Figure 1. Meridional plane of a compressor rotor.

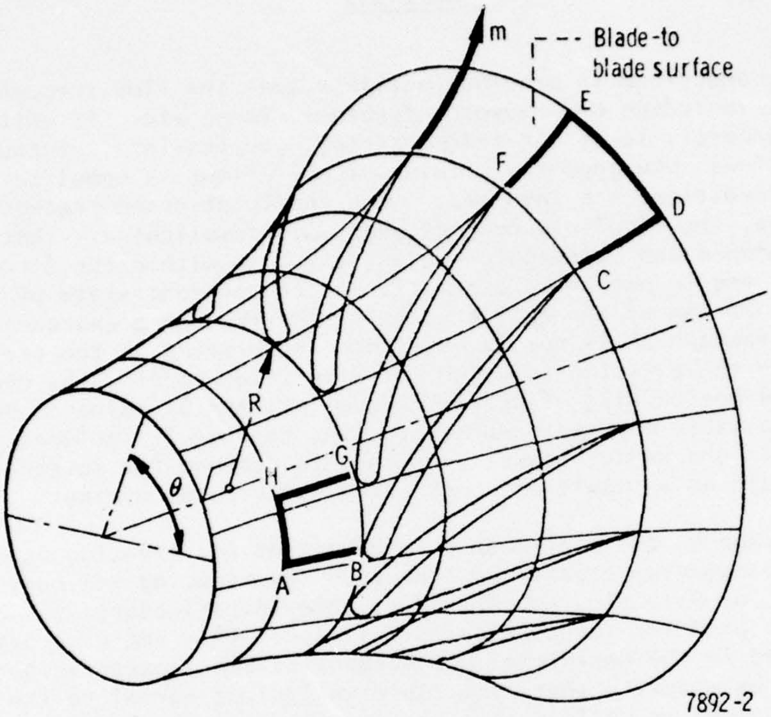


Figure 2. Blade-to-blade surface of revolution showing m-θ coordinates.

Continuity Equation

$$\frac{\partial}{\partial t} (R\Delta n\rho) = - \frac{\partial}{\partial m} (R\Delta n\rho W_m) - \frac{\partial}{\partial \theta} (\Delta n\rho W\theta) \qquad (1)$$

Meridional Momentum Equation

$$\begin{aligned} \frac{\partial}{\partial t} (R\Delta n\rho Wm) &= - \frac{\partial}{\partial m} (R\Delta n\rho W_m^2) - R\Delta nB\frac{\partial p}{\partial m} \\ &- \rho\Delta n(W_\theta - R\omega)^2 \frac{\partial R}{\partial m} - \frac{\partial}{\partial \theta} (\Delta n\rho W_m W_\theta) \qquad (2) \\ &+ \frac{R\Delta n\mu}{N_R} (\frac{4}{3} \frac{\partial^2 Wm}{\partial m^2} + \frac{1}{3R} \frac{\partial^2 W\theta}{\partial m\partial\theta} + \frac{1}{R2} \frac{\partial^2 Wm}{\partial\theta^2} \end{aligned}$$

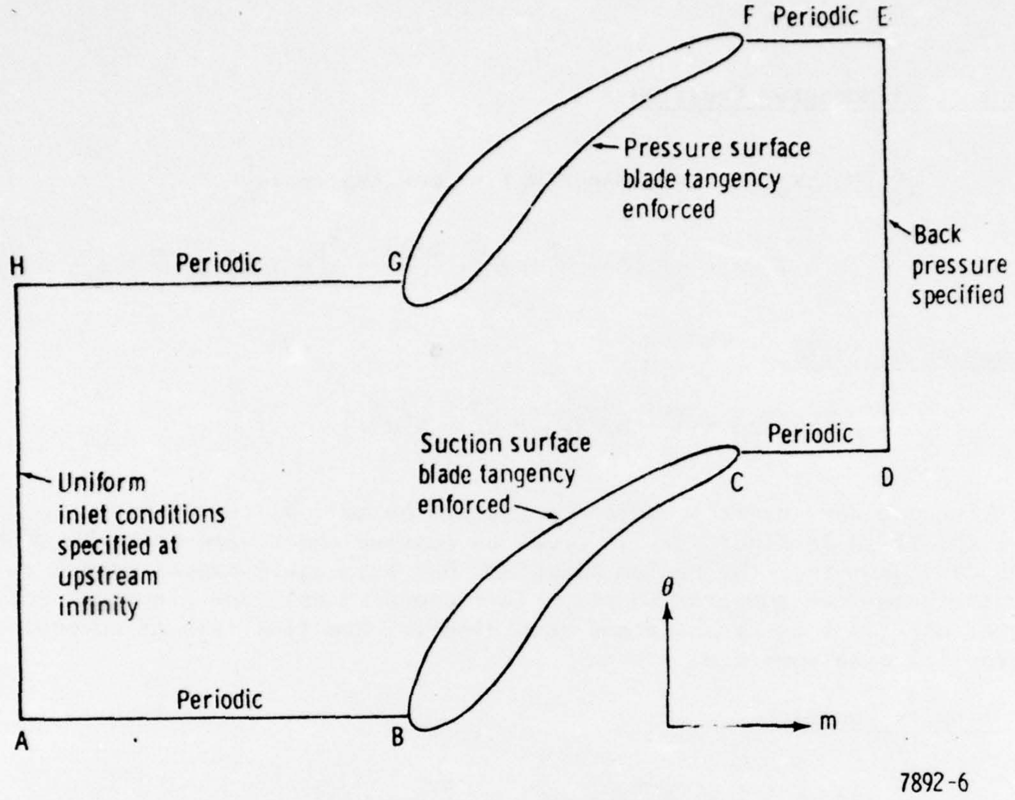

F Periodic E
Pressure surface blade tangency enforced
Back pressure specified
H
Periodic
G
Suction surface blade tangency enforced
Periodic
C
D
Uniform inlet conditions specified at upstream infinity
θ
Periodic
m
A
B
7892-6

Figure 3. Boundary conditions for a transonic compressor rotor cascade.

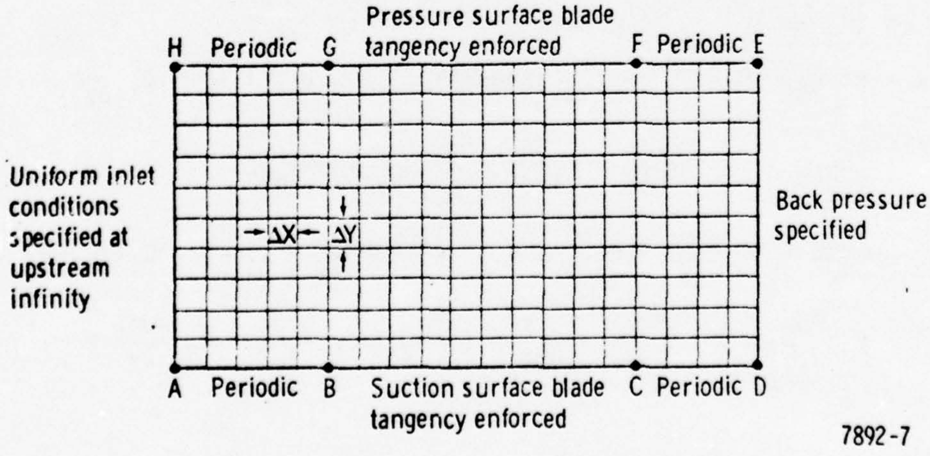

Pressure surface blade tangency enforced
H Periodic G
F Periodic E
Uniform inlet conditions specified at upstream infinity
ΔX
ΔY
Back pressure specified
A Periodic B
Suction surface blade tangency enforced
C Periodic D
7892-7

Figure 4. Transformed plane for numerical computation.

Tangential Momentum Equation

$$\frac{\partial}{\partial t}(R\Delta n\rho W_\theta) = -\frac{\partial}{\partial m}(R\Delta n\rho W_m W_\theta) - \rho\Delta n W_m (W_\theta - 2R\omega)\frac{\partial R}{\partial m} - \frac{\partial}{\partial\theta}[\Delta n(\rho W_\theta^2 + Bp)] + \frac{R\Delta n\mu}{N_R}\left(\frac{\partial^2 W_\theta}{\partial m^2} + \frac{1}{3R}\frac{\partial^2 Wm}{\partial m\partial\theta} + \frac{4}{3R^2}\frac{\partial^2 W\theta}{\partial\theta^2}\right) \tag{3}$$

Equation of State

$$p = \rho h = \rho[H_R - \frac{\gamma-1}{\gamma} M_r^2 (W_m^2 + W_\theta^2 - R^2\omega^2)] \tag{4}$$

The boundary control surface ABCDEFGHA as well as the boundary conditions are specified in Figure 3. In order to resolve the severe problems of the boundary geometry, the region ABCDEFGHA has been again mapped into a rectangle in the numerical computation plane (X-Y coordinates), see Figure 4, and the equations (1-3) are transformed accordingly. The final set of governing differential equations are:

Continuity Equation

$$\frac{\partial}{\partial t}(R\Delta n\rho) = -F'\left[\frac{\partial}{\partial x}(R\Delta n\rho W_m) - C_1\frac{\partial}{\partial y}(R\Delta n\rho W_m)\right] - \frac{1}{\theta_{21}}\frac{\partial}{\partial y}(\Delta n\rho W_\theta) \tag{5}$$

Meridional Momentum Equation

$$\begin{aligned}
&\frac{\partial}{\partial t}(R\Delta n\rho W_m) \\
&= -F'\left[\frac{\partial}{\partial x}(R\Delta n\rho W_m^2) - C_1\frac{\partial}{\partial y}(R\Delta n\rho W_m^2)\right] - R\,\Delta n\,BF'\left(\frac{\partial p}{\partial x} - C_1\frac{\partial p}{\partial y}\right) \\
&\quad + \rho\Delta n(W_\theta - R\omega)^2 F'\left(\frac{\partial R}{\partial x} - C_1\frac{\partial R}{\partial y}\right) - \frac{1}{\theta_{21}}\frac{\partial}{\partial y}(\Delta n\rho W_m W\theta) \qquad (6) \\
&\quad + \frac{R\Delta n\mu}{N_R}\left\{\frac{4}{3}F''\left(\frac{\partial W_m}{\partial x} - C_1\frac{\partial W_m}{\partial y}\right) + \frac{4}{3}(F')^2\frac{\partial^2 W_m}{\partial x^2} - C_2\frac{\partial Wm}{\partial y} + C_1\left(2C_3\frac{\partial Wm}{3\partial 2y}\right.\right. \\
&\quad \left.\left. - 2\frac{\partial^2 Wm}{\partial x\partial y} + C_1\frac{\partial^2 Wm}{\partial y^2}\right) + \frac{F'}{3R\theta_{21}}\left(\frac{\partial^2 W\theta}{\partial x\partial y} - C_1\frac{\partial^2 W\theta}{\partial y^2} - C_3\frac{\partial W\theta}{\partial y}\right) + \frac{1}{R^2\theta_{21}^2}\frac{\partial^2 Wm}{\partial y^2}\right\}
\end{aligned}$$

Tangential Momentum Equation

$$
\begin{aligned}
&\frac{\partial}{\partial t}\,(R\Delta n\rho W_\theta) \\
&= -F'\left[\frac{\partial}{\partial x}\,(R\Delta n\rho WmW_\theta) - C_1\frac{\partial}{\partial y}\,(R\Delta n\rho WmW_\theta)\right] \\
&\quad -\rho\Delta nW_m(W_\theta-2R\omega)F'\left(\frac{\partial R}{\partial x} - C_1\frac{\partial R}{\partial y}\right) - \frac{1}{\theta_{21}}\,\frac{\partial}{\partial y}\left[\Delta n(\rho W_\theta^2+Bp)\right] \qquad (7) \\
&\quad + \frac{R\Delta n\mu}{N_R}\,\left\{F''\left(\frac{\partial W_\theta}{\partial x} - C_1\frac{\partial W_\theta}{\partial y}\right) + (F')^2\,\frac{\partial^2 W_\theta}{\partial x^2} - C_2\,\frac{\partial W_\theta}{\partial y} + C_1\left(2C_3\frac{\partial W_\theta}{\partial y}\right.\right. \\
&\quad \left.- 2\,\frac{\partial^2 W}{\partial x\partial y} + C_1\frac{\partial^2 W}{\partial_y 2}\right) + \frac{F'}{3R\theta_{21}}\,\left(\frac{\partial^2 Wm}{x\ y} - C_1\frac{\partial^2 Wm}{\partial_y 2} - C_3\,\frac{\partial Wm}{\partial y}\right) + \frac{4}{3R^2\theta 2_{21}}\,\left.\frac{\partial^2 W_\theta}{\partial y^2}\right\}
\end{aligned}
$$

Equation of State

$$p = \rho h = \rho[H_R - \frac{\gamma-1}{2}\,M_r^2(W_m^2 + W_\theta^2 - R^2\omega^2)] \qquad (8)$$

It is obvious that the above set of equations has been written in the so-called "conservative" or "divergence" form with exceptions of the transport terms, and that the gas has been assumed perfect with constant viscosity.

The set of equations was solved by the MacCormack finite differencing scheme in an iterative or time-marching manner. The steady-state solution is obtained, if the difference between two successive solutions of density is less than 10^{-4} times its current value. In normal cases, it takes about 3-4 minutes to obtain the flow field solutions at 253 mesh points. For better accuracy, the computation time is much longer; therefore, there is a need to find either better numerical scheme or better mathematical model than the current one. The following section is to present a new numerical scheme for solving the same set of equations as well as boundary conditions.

METHOD OF SOLUTION

The Alternating-Direction Implicit difference approximations (ADI) have been applied to the two-dimensional Navier-Stokes equations in cylindrical or cartesian coordinates. It has been found that the conservative differencing of inviscid terms makes the method very well behaved in the presence of internal shocks. The method is also attractive for:

1. High speed flows with local regions requiring high resolution. Explicit methods require a reduced Δt when Δx is reduced due to the stability limitations, thus the computing time is increased.

2. Flows which are initially high speed and decelerated approaching to incompressible conditions at large time. Here, the explicit methods become inefficient for low speed flow.

3. Low speed flows with variable density.

In general, the implicit finite difference methods are unconditionally stable, at least for linear cases, and the solution converges much faster than the explicit finite difference schemes, because this method solves simultaneous equations, so that the effect of one point in the grid system is felt by many more points in the field than those of the explicit methods. Therefore, it is chosen to replace the MacCormack explicit difference method used in the original program.

The solution of the nonlinear partial differential equations (6) - (8) given in previous section is started by replacing the differential operators by their corresponding finite difference ones. The ADI is a two step approximation scheme. Each step the time interval of $\Delta t/2$ is used. For the odd steps derivatives in the x-direction are implicit (written in terms of dependent variables at new time point) and derivatives in y-direction are explicit (written in terms of dependent variables at old time point). For even steps, the x-direction derivatives are explicit while the y-direction derivatives are implicit. According to this idea, the governing differential equations are transformed as following.

The continuity equation is chosen as an example to show the detailed formulation procedure. In the x-y coordinates, it is given as

$$R\Delta n \frac{\partial \rho}{\partial t} = -F' \left(R\Delta n \frac{\partial \rho Wm}{\partial x} + \rho Wm \frac{\partial R\Delta n}{\partial x} - C_1 R\Delta n \frac{\partial \rho Wm}{\partial y}\right) - \frac{R\Delta n}{R\theta_{21}} \frac{\partial \rho W\theta}{\partial y} \qquad (9)$$

By expanding and rearranging, it becomes

$$\frac{\partial \rho}{\partial t} + F' \left(Wm \frac{\partial \rho}{\partial x} + \rho \frac{\partial Wm}{\partial x}\right) = -F' \left(\frac{\rho Wm}{R\Delta n} \frac{\partial R\Delta n}{\partial x} - C_1 \frac{\partial \rho Wm}{\partial y}\right) - \frac{1}{R\theta_{21}} \frac{\partial \rho W\theta}{\partial y} \qquad (10)$$

For x-direction implicit step, the finite difference transformation results:

$$\frac{1}{\Delta t/2} (\rho^{n+\frac{1}{2}} - \rho^{n}) + \frac{F'Wm}{2\Delta x} (\rho^{n+\frac{1}{2}}_{i+1} - \rho^{n+\frac{1}{2}}_{i-1}) + \frac{F'}{2\Delta x} (Wm^{n+\frac{1}{2}}_{i+1} - Wm^{n+\frac{1}{2}}_{i+1})$$

$$= \left[-F' \left(\rho WmRnx - C_1 \frac{\partial}{\partial y} \rho Wm\right) - \frac{\partial}{\partial y} (\rho W\theta)/R\theta_{21}\right]^{n} \qquad (11)$$

It may be put into the final form of (12). The formulation procedure for the momentum equations is identical. Their final results are also given below with their coefficients defined in the Appendix. The y-direction implicit step is carried out in similar fashion. Their results are not listed due to limited length of this brief report.

Continuity Equation:

A11(m,1)D(I-1,J,2)+A11(M,2)D(I,J,2)+A11(M,3)D(I+1,J,2)

+A12(M,1)WM(I-1,J,2)+A12(M,2)WM(I,J,2)+A12(M,3)WM(I+1,J,2) (12)

= B1(M)

Meriodional Momentum Equation:

A21(M,1)D(I-1,J,2)+A21(M,2)D(I,J,2)+A21(M,3)D(I+1,J,2)

+A22(M,1)WM(I-1,J,2)+A22(M,2)WM(I,J,2)+A22(M,3)WM(I+1,J,2) (13)

+A23(M,1)WT(I-1,J,2)+A23(M,2)WT(I,J,2)+A23(M,3)WT(I+1,J,2)

= B2(M)

Tangential Momentum Equation:

A31(M,1)D(I-1,J,2)+A31(M,2)D(I,J,2)+A31(M,3)D(I+1,J,2)

+A32(M,1)WM(I-1,J,2)+A32(M,2)WM(I,J,2)+A32(M,3)WM(I+1,J,2) (14)

+A33(M,1)WT(I-1,J,2)+A33(M,2)WT(I,J,2)+A33(M,3)WT(I+1,J,2)

= B3(M)

The above set of three equations are applied to each point on a constant y line (21 interior points are used for the present study, M varies from 1 to 21). The resulting set of 63 simultaneous algebraic equations are solved by the subroutine BANDEQS developed by Dr. Petty [24] of AFAPL.

CONCLUSION AND RECOMMENDATION

As stated in the Objectives section of this report that the present work is to:

1. improve, if possible, the best available computer program for simulating the transonic flow of a viscous, compressible fluid through a cascade of blades in axial flow turbomachines, and

2. explore the new mathematical models and/or solution techniques which will be more promising to handle more realistic transonic flows in the turbomachine blade elements.

To achieve the first objective, the computer program developed by Kurzrock and Novick [1], which is the program available to the AFAPL for solving the transonic flow with dominant viscous terms retained in the Navier-Stokes equations, has been studied. It was decided that the improvements may be made by (1) simplifying the governing differential equations, and (2) developing a new solution technique to replace the one used in the original program.

By order of magnitude analysis, it has been found that two more viscous terms retained in the original program, in which all second order derivative terms are kept, are of lower order of magnitude. Therefore, they are eliminated. The flowfield calculations with this modified program can obtain the same accuracy with computing time reduced up to 30% in some cases. However, the reduction in computing time is not consistent in all cases. Therefore, the decision has been made to develop a new subroutine for flowfield solutions.

ADI difference scheme was chosen for reasons outlined in the Method of Solution section. The formulation and programming have been completed. The debugging process for the FINITE (ADI) subroutine has also been completed. The difficulties of making the subroutine FINITE (ADI), an implicit scheme, and the rest of subroutines of the complete program, most of them being explicit schemes, compatible have been numerous and subtle. Unfortunately, there is no time left to complete this part of the research program.

According to a rough estimation of computation time required to obtain a converged solution of the flowfield by the method of ADI, it has been found that the reduction in computing time, if possible, may not be much. Although it usually takes less iterations for an implicit scheme to converge than that of explicit schemes, it takes much longer time to carry out each iteration because the solution of a large number of simultaneous equations is necessary in applying the implicit schemes, while the solution of explicit scheme only requires simple arithmetic evaluations. Furthermore, when higher accuracy is needed, the solution of a larger number of simultaneous equations may not only exceed the capacity of the computer storage, but also makes the computing time too long to be economical. It has been found from several trial runs that even by using the BANDEQS, a subroutine developed by Petty [24] for efficiently computing the solution of band matrix equations, it takes about one second of of CDC6600 execution time to calculate one integration cycle, which is about five times as long as it takes to complete one integration step by the MacCormack explicit differencing scheme. Therefore, if by using the ADI the number of integration steps would be cut down from 1000 to 200; there would not be any savings of computing time.

In addition, the mathematical model solved by Kurzrock and Novick [1] has some drawbacks. First of all, the location of the stream surfaces is assumed to be known a priori. For axial-flow compressor applications, these locations are usually based on a three-dimensional equilibrium analysis, and are not modified throughout the solution of the Navier-Stokes equation. Secondly, flow field variations normal to the stream surfaces are ignored. This assumption is not true in most cases. Thirdly, the assumptions of a linear relation between the inlet and outlet locations of the stream surfaces, the radius (R) and the stream surface convergence, Δn, being known functions of the meridional distance (m), the fluid being perfect gas with constant viscosity and thermal conductivity, and so on. Beside the above idealizing and/or simplifying assumptions, the convergence of the solution is highly dependent on the initial guessing of the flow field properties, difficulties in obtaining solutions often occur at the leading and trailing edges, and the governing nonlinear differential equations become more and more complicated through the successive transformations in order to put the final equation in a convenient computation domain. All these drawbacks lead us to believe that a different approach is needed.

The Finite Element Method has the advantages of generality and simplicity of both mathematical formulation and numerical solution. It is well-suited for solving problems with irregular boundary shapes including the curved boundaries as well as involved boundary conditions, so that not only the problems occurring at the leading and trailing edges of the finite difference approaches may be eliminated, but also the successive coordinate transformations are not needed. Time and space discretigation can be distinguished easily. Each portion of the formulation can be studied separately and causes of accuracy and stability problems can be investigated. In some cases, the combination of simple finite difference processes in time with finite element space discretization has also proved to be very effective.

In short, it becomes obvious that the finite element method is very promising in handling the transonic flows of a viscous and compressible fluid through the blade elements of turbomachines. Additional efforts in the advancement of its application to the turbomachine flowfield simulation is recommended.

REFERENCES

1. Kurzrock, J.W. & Novick, A.S., "Transonic Flow Around Compressor Blade Elements", Vol. I & II, T.R. AFAPL-TR-73-69 (1973)

2. Merchant, W. & Collar, A.R., "Flow of an Ideal Fluid Past a Cascade of Blades," Aero Research Council R. and M. No. 1893 (1941)

3. Gostellow, J.P., "Potential Flow through Cascades, Extensions to an Exact Theory," Aero. Research Council Current Paper 808 (1964)

4. Schlichting H. & Scholz, N., "Uber die Theoretische Berechnung der Stromungsverluste Eines Ebenen Schaufelgitters," Ing.-Arch., Bd XIX, Heft 1 (1951)

5. Martensen E., Calculation of Pressure Distribution Over Profiles in Cascade in Two Dimensional Potential Flow by Means of a Fredholm Integral Equation, Arch. for Rat. Mech. and Anal., Vol 3, No. 3 pp 325 (1959)

6. Wilkinson, D.H., "A Numerical Solution of the Analysis and Design Problems for the Flow Past One or More Aerofoils in Cascade," Aero. Research Council, R. & M. 3545 (1968)

7. Von, Karman T., "Compressibility Effects in Aerodynamics," Journal Aero. Sci. Vol 8, No. 9 (1941)

8. Tsien, H.S., "Two Dimensional Subsonic Flow of Compressible Fluids," Journal Aero. Sci. Vol 6, No. 10 (1939)

9. Imbach, H.E., "Calculation of the Compressible Potential Flow Around Profiles in Cascade," Brown Boveri Review, Vol 51, No. 12, pp 752 (1964)

10. Price, D., "Calculation of the Velocity Distribution Around Profiles in Cascade, in Two-Dimensional Potential Flow. (unpublished Rolls-Royce report) 1963.

11. Bindon, J.P. and Carmichael, A.D., "Streamline Curvature Analysis of Compressible and High Mach Number Cascade Flows," J. Mech. Eng. Sci. Vol 13, No. 5, pp 344 - 357 (1971)

12. Katsanis, T., "Use of Arbitrary Quasi Orthogonals for Calculating the Flow Distribution on a Blade to Blade Surface in a Turbomachine," NASA TN-D-2809 (1965)

13. Wilkinson, D.H., Stability, Convergence and Accuracy of Two-Dimensional Streamline Curvature Methods Using Quasi-Orthogonals," Proc. I. Mech. E., Vol 184, Pt. 3G, pp 108-119 (1970)

14. Smith, D.J.L., and Frost, D.H., "Calculation of the Flow Past Turbomachines Blades," Proc. I. Mech. E., Vol 184 Pt. 3G, pp 219-233 (1970)

15. Katsanis, T., "Computer Program for Calculating Velocities and Streamlines on a Blade-to-Blade Stream Surface of a Turbomachine," NASA TN D-4525 (1968)

16. Perkins, H.J., "A Note on the Through-Flow Analyses of Turbomachines," Aero. Research Council, Report No. 34, 315 (1973).

17. Hobson, D., "A Hodograph Method for the Design of Transonic Turbine Blades," Aero. Research Council Report No. 34, 131

18. Curtis, E.M., Hutton, M.F., and Wilkinson, D.H., "Theoretical and Experimental Work on Two-Dimensional Turbine Cascades with Supersonic Outlet Flow," Proc. I. Mech. E. Warwick Conference (1973)

19. McDonald, P.W., "The Computation of Transonic Flow through Two-Dimensional Gas Turbine Cascades," Am. Soc. Mech. Engrs. Paper 71-GT-89 (1971).

20. Gopalakrishnan, S. and Bozzola, R., "A Numerical Technique for the Calculation of Transonic Flows in Turbomachinery Cascades," Am. Soc. Mech. Engrs. Paper 71-GT-42 (1971).

21. Marsh H., and Merryweather H., "The Calculation of Subsonic and Supersonic Flows in Nozzles. Proc. I. Mech. E., Salford Conference on Internal Flow Paper 22 (1971)

22. Daneshyar H. and Glynn, D., "The Calculation of Flow in Nozzles Using a Time-Marching Technique Based on the Method of Characteristics," Int. J. Mech. Sci. (In press 1973)

23. Thompson D.S., "Finite Element Analysis of the Flow Through a Cascade of Aerofoils," Aero. Research Council Report No. 34, 412 (1973)

24. Petty, J.S., "Research in Computational Fluid Dynamics," Tech Report AFARL 75-0209, in print (1975)

APPENDIX

The list of coefficients used in the ADI finite difference equations (x-direction implicit step)

$A11(M,1)=-A11(M,3)=-F'Wm\Delta t/4\Delta x$

$A11(M,2)=1$

$A12(M,1)=-A12(M,3)=-F'\rho\Delta t/4\Delta x$

$A12(M,2)=0$

$A13(M,1)=A13(M,2)=A13(M,3)=0$

$B1(M)=\rho-(F'(R_{nx}\rho Wm-C_1\frac{\partial}{\partial y}\rho Wm)+\frac{\partial}{\partial y}(\rho W\theta)/R\theta_{21})$

$A21(M,1)=-A21(M,3)=-F'(Wm^2+Bh)\ \Delta t/4\Delta x$

$A21(M,2)=Wm$

$A22(M,1)=-A22A-A22B$

$A22(M,3)=A22A-A22B$

$A22(M,2)=\rho-2A22B$

$A22A=(F'\rho Wm(\gamma+1)/\gamma-F''4\mu/3N_R)\Delta t/4\Delta x$

$A22B=(F')^2 2\mu\Delta t/3N_R(\Delta x)^2$

$A23(M,1)=-A23(M,3)=F'\rho W\theta\Delta t(\gamma-1)/\gamma 4\Delta x$

$A23(M,2)=0$

$$\begin{aligned}
B2(M)=2\rho Wm+(F'(C_1\ \frac{\partial}{\partial y}\ (\rho W_m^2)-\rho W_m^2 Rnx)\ +BC_1F'\ \frac{\partial p}{\partial y} \\
+\rho(W\theta-R\omega)^2\ F'Rx-\frac{\partial}{\partial y}\ (\rho WmW\theta)/R\theta_{21}\ +\ (\mu/N_R)(\frac{\partial Wm}{\partial y}\ 4/3 \\
((F')^2\ (2C_1C_3-C_2)-C_1F'')\ +\ \frac{\partial^2 Wm}{\partial y^2}\ (C_1^2(F')^2\ 4/3\ +\ (1/R\theta_{21})^2) \\
-\ \frac{\partial^2 Wm}{\partial x\partial y}\ (F')^2C_1\ \ 8/3+F'\ (\frac{\partial^2 W\theta}{\partial x\partial y}\ -C_1\ \frac{\partial^2 W\theta}{\partial y^2}\ -\ C_3\ \frac{\partial W\theta}{\partial y}\ )))\Delta t/2
\end{aligned}$$

$A32(M,1)=-A31(M,3)=-F'WmW\theta\Delta t/4\Delta x$

$A31(M,2)=W\theta$

$$A32(M,1)=A32(M,3)=-F'\rho W\theta\Delta t/4\Delta x$$

$$A32(M,2)=0$$

$$A33(M,1)=A33A-A33B$$

$$A33(M,3)=A33A+A33B$$

$$A33(M,2)=\rho-2A33A$$

$$A33A=-(F')^{2}\mu\Delta t/2(\Delta x)^{2}N_{R}$$

$$A33B=(F'\rho Wm-F'\mu/N_{R})\Delta t/4\Delta x$$

$$B3(M)=2\rho W\theta+(F'(C_{1}\frac{\partial}{\partial y}(\rho WmW\theta)-\rho WmW\theta Rnx)-Wm(W\theta-2R\omega)F'Rx$$

$$-B\frac{\partial p}{\partial y}-\frac{\partial}{\partial y}(\rho W_{\theta}^{2})+((F')^{2}(2C_{1}C_{3}-C_{2})\frac{\partial W\theta}{\partial y}-F''C_{1}\frac{\partial W\theta}{\partial y}$$

$$+F'(\frac{\partial^{2}Wm}{\partial x\partial y}-C_{1}\frac{\partial^{2}Wm}{\partial_{y}2}-C_{3}\frac{\partial Wm}{\partial y})/3R\theta_{21})\mu/N_{R})\Delta t/2$$

where: $F'=\frac{dF}{dm}$

$$F''=\frac{d^{2}F}{dm^{2}}$$

$$Rx=\frac{\partial R}{\partial x}/R\Delta n$$

$$Rnx=\frac{\partial R\Delta n}{\partial x}/R\Delta n$$

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT-PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

AN ANALYSIS OF VARYING MATERIAL PROPERTIES:
A MEASURE OF DAMAGE

Prepared by:	Donald C. Stouffer, Phd.
Academic Rank:	Associate Professor
Department and University:	Department of Engineering Analysis University of Cincinnati
Assignment:	
(Laboratory)	Materials Laboratory
(Division)	Metals and Ceramics
(Branch)	Metals Behavior
USAF Research Colleague:	T. Nicholas
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

AN ANALYSIS OF VARYING MATERIAL PROPERTIES:

A MEASURE OF DAMAGE

by

Donald C. Stouffer

ABSTRACT

A measure of damage is developed that is based entirely on the changes of the physical properties of a material that are characterized by changes in the constitutive law. These changes are defined mathematically by Frechét differentiation of the material functional with respect to the deformation history. The tensorial measure of damage is initially defined for a simple material and later generalized to any constitutive property. Several other properties of the damage measure are developed including two presentations for the measure, the coupling relationship between different independent variables such as thermal and mechanical histories, the effect of residual stress distributions, the effect of auxillary sensing devices, and the relationship to thermodynamic dissipation and ideal materials. It is also pointed out how the damage measure can be used in the development of constitutive equations. The concept of damage given here is completely new and represents a significant departure from the traditional approach.

INTRODUCTION

The importance of knowing the state of damage in a material is of unquestionable value since it leads directly to a method of predicting failure. The most popular methods of estimating damage are based on measures of time-to-failure or cycles-to-failure (See [1] for example), however, the value of these measures as fundamental criteria is questionable since neither is a basic material parameter. Recently, Berkovits [2] used the accumulated inelastic strain as a measure of the damage in analyzing the high temperature behavior of metals. However this measure excludes the effect of elastic cycling and reversals in the damage process such as annealing, that can be present after a plastic deformation at a high temperature.

Constitutive changes such as annealing demonstrate that the concept of damage is much more general than just trying to predict the degradation of a material. The formulation should also allow for the restoration or even the enhancement of material properties. In fact, the classification of "damage" or "enhancement" depends upon the particular use of the material. For example, a change in a mechanical property that is harmful in one situation may be acceptable or even desirable in another situation.

To demonstrate the basic ideas in this approach to damage, consider an experimental procedure for two initially identical metal test samples. Assume that sample one is stretched or twisted in such a way that the material properties have been changed but so that the outward appearance looks the same. Then assume an experimentalist subjects both test samples to identical deformation histories. The response for test samples one and two would in general be different. This difference in response would reflect the change in mechanical properties produced by the previous deformation history on test sample one. It is this change or variation in material response that can be associated with damage.

In this analysis a measure of damage is developed based entirely on the constitutive properties of the material through the constitutive law. Recall that a mechanical constitutive equation is used to characterize the physical properties of the material at each particular time and position in the deformation process. Thus, it is possible to take the constitutive functional as characterizing the mechanical material state. Since the change in material properties is identified with the change in material state, these state changes can be measured directly by changes in response predicted by the constitutive functional.

The formulation of this damage measure leads directly to a derivative operation of the material functional. Thus many of the properties of functional calculus can be used to establish various properties of the damage measure. It is worth noting that the accuracy of the damage measure depends upon the sensitivity of the constitutive relationship. However, the accuracy of a constitutive equation can be improved, since the changes of the constitutive response can be associated with the deformation history producing the variation in material properties through a derivative operation of the material functional.

DAMAGE IN A SIMPLE MATERIAL

Let R be a region in a three dimensional Cartesian space. The position of a generic particle in a continuous medium in a reference configuration is given by the material coordinates $\underset{\sim}{\bar{X}}$ in R and the position of the particle in a deformed configuration at time t is given by the spatial coordinates $\underset{\sim}{x}$ also in R. Assume the deformation process is smooth and continuous so that a unique coordinate mapping and its inverse exist; *i.e.*,

$$\underset{\sim}{x} = \underset{\sim}{x}(\underset{\sim}{\bar{X}},t) \text{ and } \underset{\sim}{\bar{X}} = \underset{\sim}{\bar{X}}(\underset{\sim}{x},t) \qquad (2.1)$$

for t $\varepsilon(-\infty,\infty)$. The gradient, $\underset{\sim}{F}$, of the mapping $\underset{\sim}{x}(\underset{\sim}{\bar{X}},t)$ is given by

$$\underset{\sim}{F}(\underset{\sim}{\bar{X}},T) = \nabla_{\underset{\sim}{\bar{X}}}\underset{\sim}{x}(\underset{\sim}{\bar{X}},t) \qquad (2.2)$$

and designated as the deformation gradient relative to the reference configuration. The velocity of a generic particle $\underset{\sim}{x}(\underset{\sim}{\bar{X}},t)$ is given by the material time derivative of the mapping $\dot{\underset{\sim}{x}}(\underset{\sim}{\bar{X}},t)$ and the velocity gradient $\underset{\sim}{L}$ is given by

$$\underset{\sim}{L}(\underset{\sim}{\bar{X}},t) = \nabla_{\underset{\sim}{\bar{X}}}\dot{\underset{\sim}{x}}(\underset{\sim}{\bar{X}},t). \qquad (2.3)$$

Further, the relationships

$$\dot{\underset{\sim}{F}} = \underset{\sim}{L}\,\underset{\sim}{F} \text{ and } \underset{\sim}{L} = \dot{\underset{\sim}{F}}\,\underset{\sim}{F}^{-1} \qquad (2.4)$$

exist assuming that $|\underset{\sim}{F}|\neq 0$. The history $\underset{\sim}{h}^t(s)$ up to time t of any scaler or tensor function $\underset{\sim}{h}(t)$ is defined as

$$\underset{\sim}{h}^t(s) = \underset{\sim}{h}(t-s) \qquad (2.5)$$

for S $\varepsilon[0,\infty]$ noting that $\underset{\sim}{h} = \underset{\sim}{h}(t) = \underset{\sim}{h}^t(0)$.

Consider now a mechanical constitutive equation for a simple material as given by Truesdell and Noll [3]. Here it is assumed that the Cauchy stress, $\underset{\sim}{T}(\underset{\sim}{x},t)$, at any time t and position $\underset{\sim}{x}$ in R is given by a functional of the deformation gradient alone and that this relationship must be of the form

$$\underset{\sim}{T}(\underset{\sim}{x},t) = p\,\underset{\sim}{F}\,\underset{\sim}{\overset{\infty}{\underset{s=0}{Q}}}\,[\underset{\sim}{F}^t(s)]. \qquad (2.6)$$

The constitutive equation (2.6) satisfies the Principle of Material Objectivity and is consistent with the principles of thermodynamics, Coleman [4]. The functional $\underset{\sim}{Q}$ is assumed to characterize all of the relevant mechanical properties, i.e., it is taken to accurately reflect the macroscopic properties of the material at any position $\underset{\sim}{x}$ at time t in a deformed configuration of the continuim.

The first Piola - Kirchhoff Stress Tensor, $\underset{\sim}{S}$, is related to the Cauchy stress by

$$\underset{\sim}{S}(\underset{\sim}{x},t) = \frac{1}{p}\underset{\sim}{F}^{-1}\underset{\sim}{T} = \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}[\underset{\sim}{F}^{t}(s)]. \tag{2.7}$$

and follows directly from equation (2.6). The constitutive equation (2.7) can now be used to develop a representation for the change in the material properties.

In order to measure the variation in the material properties using $\underset{\sim}{Q}$, it is necessary to separate the direct effect of the deformation from the intrinsic material properties characterized by the functional $\underset{\sim}{Q}$. To accomplish this let $\underset{\sim}{R}^{t}(s)$ represent some reference deformation gradient history applied at time t=0 with the properties that $\underset{\sim}{R}^{t}(s)=0\ \forall\ s\ \varepsilon[t,\infty]$ and $\underset{\sim}{R}^{t}(s)\neq\underset{\sim}{0}\ \forall\ s\ \varepsilon[0,t]$. The stress $\underset{\sim}{S}_{R}$ at time t resulting from $\underset{\sim}{R}^{t}(s)$ can be determined from equation (2.7) as

$$\underset{\sim}{S}_{R}(\underset{\sim}{x},t) = \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}[\underset{\sim}{R}^{t}(s)]. \tag{2.8}$$

Next denote a second deformation gradient history $\underset{\sim}{F}_{d}^{t}(s)$ with the property that $\underset{\sim}{F}_{d}^{t}(s)$ is aribitrary for s $\varepsilon[t,\infty]$ and such that the stress

$$\underset{\sim}{S}_{d}(t) = \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}[\underset{\sim}{F}_{d}^{t}(s)] = 0 \text{ for } t > 0. \tag{2.9}$$

This loading sequence corresponds to a load and recovery experiment.

Next let us define a deformation history $\underset{\sim}{F}_{D}^{t}(s)$ resulting from the superposition of $\underset{\sim}{R}^{t}(s)$ on $\underset{\sim}{F}_{d}^{t}(s)$ with the origin of adjusted to zero at time t; that is, let

$$\underset{\sim}{F}_{D}^{t}(s) = {}^{*}\underset{\sim}{F}_{d}^{t}(s) + \underset{\sim}{R}^{t}(s) \tag{2.10}$$

for all s $\varepsilon[0,\infty]$, see Figure 1. The quantity ${}^{*}\underset{\sim}{F}_{d}^{t}(s) = \underset{\sim}{F}_{d}^{t}(s) - \underset{\sim}{F}(t)$ is the difference history and represents a measure of the deformation relative to the current configuration.

The stress $\underset{\sim}{S}_{D}(t)$ due to the deformation history $\underset{\sim}{F}_{D}^{t}(s)$ can be determined from equation (2.7) as

$$\underset{\sim}{S}_{D}(t) = \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}[\underset{\sim}{F}_{D}^{t}(s)]. \tag{2.11}$$

The stress $\underset{\sim}{S}_{D}(\tau)\ \tau\ \varepsilon[0,t]$ is, in general, different from the stress $\underset{\sim}{S}_{R}(t)$ due to the prior deformation history ${}^{*}\underset{\sim}{F}_{d}^{t}(s)$. The difference in the response as observed by $\underset{\sim}{S}_{R}(t)$ and $\underset{\sim}{S}_{D}(t)$ for t $\varepsilon[0,t]$ is defined as the MATERIAL VARIATION TENSOR $\underset{\sim}{V}(t)$; i.e.,

$$V(t) \equiv S_D(t) - S_R(t). \tag{2.12}$$

The tensor $V(t)$ is a measure of the relative change in the material properties due to the predeformation $*F_d^t(s)$. Observe that the deformation $F_d^t(s)$ does not contribute directly to variation $V(t)$ (since $S_d(t)=0$ for $t>0$) but only through history effects which are manifested by changes in the material microstructure. Observe that the material variation V is a tensor valued functional. This reflects the fact that deformation histories in different directions can produce different types of changes in the material microstructure. For example, constitutive changes resulting from plastic tensile and shear deformations in two metal test samples would in general yield different microstructures in the deformed configuration.

The material variation tensor $V(t)$ can be determined for any set of deformation histories $R^t(s)$ and $*F_d^t(s)$ once a specific constitutive representation is specified. A representation for the material variation tensor V can be developed in terms of the constitutive functional Q by substituting equation (2.8) and (2.11) into (2.12) to obtain

$$V(t) = \underset{s=0}{\overset{\infty}{Q}} [R^t(s) + *F_d^t(s)] - \underset{s=0}{\overset{\infty}{Q}} [R^t(s)]. \tag{2.13}$$

Observed that $*F_d^t(s)$ is the deviation history from the reference history R^t at any $t \geq 0$.

The material variation tensor is formulated on the basis that the stress $S_d(t)$ calculated from Equation (9) is zero for all $t>0$. This may not be possible if a residual stress is present as resulting from a nonhomogeneous deformation field. The residual stress $S_{RES}(t,x)$ at time t and position x corresponding to a stress free boundary condition can be written as

$$S_{RES}(t,x) = \underset{s=0}{Q} [F_d^t(s,x)] \tag{2.14}$$

recalling that $F_d^t(s,x)$ is the "preworking" deformation history. Since the residual stress distribution is elastic, the corresponding deformation, $F_{RES}(t,x)$ that can be recovered by sectioning the body, can be totally determined from the elastic properties of the body at time t and position x. Let us assume the elastic properties of the body are completely characterized by the function f such that the residual deformation F_{RES} is given by

$$F_{RES}(t,x) = f[S_{RES}(t,x),x,t]. \tag{2.15}$$

The difference history, $F_d^t(s)$, corresponding to a stress free state for $t>0$ is given by

$$*F_d^t(s) = F_d^t(s) - \{F(t) - F_{RES}(t)\} \tag{2.16}$$

with the origin adjusted to zero at time t. Hence on using Equations (2.16) in (2.13) the material variation can be determined in the presence of a residual stress distribution.

Two useful scalar measures of the variation in material properties are the magnitude of the variation at any time τ, given by

$$||\underset{\sim}{V}(\tau)|| = [\underset{\sim}{V}(\tau) \cdot \underset{\sim}{V}(\tau)]^{\frac{1}{2}} \tag{2.17}$$

which is the distance between the stress paths $\underset{\sim}{S}_D(t)$ and $\underset{\sim}{S}_R(t)$ in stress space. The total accumulated variation in material properties upto time t is

$$A_V = \int_o^t ||\underset{\sim}{V}(\tau)||dt. \tag{2.18}$$

The measure A_V is the area between the stress paths in stress space.

PROPERTIES OF THE MATERIAL VARIATION TENSOR

a) Fading Memory

The damage or enhancement of the mechanical properties as measured by the material variation $\underset{\sim}{V}(t)$ has the representation demonstrated in euqation (2.8) that is similar in structure to the Fréchet derivative Q. This property can be used to develop a representation for $\underset{\sim}{V}$ and establish various properties of $\underset{\sim}{V}$. The present theory can be put on a firm physical and theoretical basis by assuming the material has the property that the memory of the material fades in time. This implies that deformations in the distant past has less influence on the current value of the stress than deformations in the recent past. This assumption was characterized by Coleman [4] and is briefly restated here for application in the current development.

Define the memory influence function h(s) to be positive, monotone decreasing, continuous function of s such that

$$\lim_{s\to\infty} s^{\frac{1}{2}+\delta} h(s) = 0 \tag{3.1}$$

monotonically for large s and some small $\delta>0$. The Hilbert space norm $||\underset{\sim}{F}_A(\ )||_h$ for the function $F_A(\ )$ is given by

$$||\underset{\sim}{F}_A(\ )||_h = [\int_o^\infty ||\underset{\sim}{F}_A(s)||^2 h(s)^2 ds]^{\frac{1}{2}} \tag{3.2}$$

provided $||\underset{\sim}{F}_A(\ )||_h$ is finite. The Principle of Fading Memory can now be expressed by assuming every pair of functions $\underset{\sim}{F}_d^t(s)$ and $\underset{\sim}{R}^t(s)$ have for there common domain the neighborhood of the history $\underset{\sim}{0}$ and are Fréchet-differentiable in that neighborhood with respect to the norm $||\underset{\sim}{F}_d(\ )||_h$; that is, assume

$$\underset{s=0}{\overset{\infty}{\underset{\sim}{Q}}} [\underset{\sim}{R}^t(s) + *\underset{\sim}{F}_d^t(s)] = \underset{s=0}{\overset{\infty}{\underset{\sim}{Q}}} [\underset{\sim}{R}^t(s)] + \delta \underset{s=0}{\overset{\infty}{\underset{\sim}{Q}}} [\underset{\sim}{R}^t(s) \mid *F_d^t(s)] + o(||*F_d(\ )||_h) \tag{3.3}$$

where $\delta \underset{\sim}{Q}$ is linear in $*F^t_d(s)$ and continuous in both histories.

The Fréchet differential $\delta \underset{\sim}{Q}$ can further be expressed as (see [5])

$$\delta \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}} [\underset{\sim}{R}^t(s) | *\underset{\sim}{F}^t_d(s)] = \int_o^\infty \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}{}^{(1)}[\underset{\sim}{R}^t(s);\xi] \cdot *\underset{\sim}{F}^t_d(\xi)d\xi \qquad (3.4)$$

where $\underset{\sim}{Q}^{(1)}$ is the derivative of the functional Q with respect to $R^t(s)$ at point ξ on the interval $[-\infty,t]$. The integral representation follows directly from the linearity and continuity properties of $\delta \underset{\sim}{Q}$ and $*F^t_d(s)$. Thus δQ is a linear approximation of the change in value of $\underset{\sim}{Q}$ due to changing histories from $R^t(s)$ to $R^t(s) + *F^t(s)$ and the approximation becomes exact as $*F^t(s) \to \underset{\sim}{0}$ since $o(||*\underset{\sim}{F}^t(\ )||_h) \to 0$.

Derivatives of higher order are given by

$$\delta \underset{\sim}{Q}^{(m)} = \int_o^a \int_o^a \ldots \int_o^a \overset{\infty}{\underset{\sim}{Q}}{}^{(m)}[\underset{\sim}{R}^t(s);\xi_1,\xi_2,\ldots\xi_m]\cdot *\underset{\sim}{F}^t_d(\xi_1)\cdot *\underset{\sim}{F}^t_d(\xi_2)\ldots *\underset{\sim}{F}^t_d(\xi_m)d\xi_1 d\xi_2\ldots d\xi_m \qquad (3.5)$$

where $\underset{\sim}{Q}^{(m)}$ is the mth derivative of Q at the points $\xi_1,\xi_2,\ldots,\xi_m$. Further, Q is continuous and symmetric in the m parameters $\xi_1,\xi_2,\ldots,\xi_m$. For convenience in notation, let us agree to write $\delta \underset{\sim}{Q}^{(1)} = \delta \underset{\sim}{Q}$ as indicated in (3.4).

b) Representation for $\underset{\sim}{V}$

Direct application of the extension of Taylor's theorem to functionals (Volteria, §3, [5]) can be used to determine the material variation tensor. The Taylor expansion of $\underset{\sim}{V}$ about the history $R^t(s)$ is given by

$$\underset{\sim}{V}(t) = \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}} [\underset{\sim}{R}^t(s) + *\underset{\sim}{F}^t_d(s)] - \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}} [\underset{\sim}{R}^t(s)]$$

$$= \sum_{m=1} \frac{1}{m!} \int_o^\infty \int_o^\infty \ldots \int_o^\infty \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}{}^{(m)}[\underset{\sim}{R}^t(s);\xi_1,\xi_2,\ldots\xi_m]\cdot$$

$$*\underset{\sim}{F}^t_d(\xi_1):*\underset{\sim}{F}^t_d(\xi_2):\ldots:*\underset{\sim}{F}^t_d(\xi_m)d\xi_1 d\xi_2\ldots d\xi_m \qquad (3.6)$$

provided the limit of the last term approaches zero as $n\to\infty$. Comparing equation (3.3), and (3.6) it follows that

$$o(||F_f(\ )||_h) = \sum_{m=2} \frac{1}{m!} \int_o^\infty \int_o^\infty \int_o^\infty \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}{}^{(m)}[\underset{\sim}{R}^t(s);\xi_1,\xi_2\ldots\xi_m]:$$

$$:*\underset{\sim}{F}^t_d(\xi_1):*\underset{\sim}{F}^t_d(\xi_2)\ldots:*\underset{\sim}{F}^t_d(\xi_m)d\xi_1 d\xi_2\ldots d\xi_m. \qquad (3.7)$$

Since in general the history $*F^t_d(s)$ is not near $R^t(s)$ then $o(||F_d(\ )||_h)$ was not small and the higher order integral terms of (3.7) are important.

An alternative formulation for the material variation tensor, $\underset{\sim}{V}(t)$, can be developed by observing for any time $t>0$ that

$$\underset{\sim}{V}(t) = \int_o^t \dot{\underset{\sim}{V}}(\tau)d\tau \tag{3.8}$$

since $\underset{\sim}{V}(0) = 0$. The material variation rate $\dot{\underset{\sim}{V}}$ can be written as

$$\dot{\underset{\sim}{V}}(\tau) = \frac{d}{dt} \{\overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}} [\underset{\sim}{F}{}_D^t(s)] - \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}} [\underset{\sim}{R}^t(s)]\} \tag{3.9}$$

$$= \delta\overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}} [\underset{\sim}{F}{}_D^\tau(s)|\dot{\underset{\sim}{F}}{}_D^\tau(s)] - \delta\overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}} [\underset{\sim}{R}^\tau(s)|\dot{\underset{\sim}{R}}{}^\tau(s)].$$

The derivatives $\dot{F}_D^t(s)$ and $\dot{\underset{\sim}{R}}{}^t(s)$ can be evaluated in terms history parameter s for use in equation (3.7). Recalling equation (2.10) permits

$$\dot{\underset{\sim}{F}}{}_D^t(s) = \frac{d}{dt} [\underset{\sim}{R}(t-s) + {}^*\underset{\sim}{F}_d(t-s)] = - \frac{d}{ds} [\underset{\sim}{R}(t-s) + {}^*\underset{\sim}{F}_d(t-s)]$$

$$= - \frac{d}{ds} [\underset{\sim}{R}^t(s) + \underset{\sim}{F}{}_d^t(s)], \quad \text{for } s\leq 0<\infty. \tag{3.10}$$

Using equation (3.4) and (3.10) in (3.8) gives

$$\dot{V}(t) = \int_o^\infty \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}{}^{(1)}[\underset{\sim}{F}{}_D^t(s);\xi]\cdot\{-\frac{d}{d\xi}[\underset{\sim}{R}^t(\xi)+{}^*\underset{\sim}{F}{}_d^t(\xi)]\}d\xi-\int_o^\infty \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}{}^{(1)}[\underset{\sim}{R}^t(s);\xi]\cdot\{-\frac{d}{d\xi} \underset{\sim}{R}^t(\xi)\}d\xi$$

$$=-\int_o^\infty \overset{\infty}{\underset{\sim}{Q}}{}^{(1)}[\underset{\sim}{F}{}_D^t(s);\xi]\cdot\frac{d{}^*F_d^t(\xi)}{d\xi} d\xi-\int_o^t\{ \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}{}^{(1)}[\underset{\sim}{F}{}_D^t(s);\xi]- \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}{}^{(1)}[\underset{\sim}{R}^t(s);\xi]\}\cdot\frac{dR^t(\xi)}{d\xi} d\xi \tag{3.11}$$

since $\frac{d}{d\xi} \underset{\sim}{R}^t(\xi)=0$ for $\xi\varepsilon[t,\infty]$.

Substituting equation (3.11) into (3.8) gives

$$\underset{\sim}{V}(t) = \int_{\tau=o}^t \int_{\xi=o}^\tau \{ \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}{}^{(1)}[\underset{\sim}{R}^\tau(s);\xi]- \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}{}^{(1)}[\underset{\sim}{F}{}_D^\tau(s);\xi]\}\cdot\frac{d}{d\xi} \underset{\sim}{R}^\tau(\xi)\, d\xi d\tau$$

$$- \int_{\tau=o}^t \int_{\xi=o}^\infty \overset{\infty}{\underset{s=0}{\underset{\sim}{Q}}}{}^{(1)}[\underset{\sim}{F}{}_D^\tau(s);\xi]\cdot\frac{d}{d\xi} {}^*\underset{\sim}{F}{}_d^\tau(\xi)d\xi d\tau \tag{3.12}$$

where $\underset{\sim}{V}(t)$ is the damage or enhancement in the mechanical properties due to the difference history ${}^*\underset{\sim}{F}{}_d^t(s)$ measured relative to a reference deformation $\underset{\sim}{R}^t(s)$.

c) Perfect Materials

The concept of an ideal material can now be introduced. Here, a perfect material is defined as a material that does not demonstrate changes in the material variation tensor. Notice that setting $\underset{\sim}{V}(t)=0$ is not sufficient to guarantee that the material properties are invariant in time since the integral in equation (3.8) could vanish if $\dot{\underset{\sim}{V}}$ is positive and on some subinterval of [0,t] and negative on the remaining subinterval. This corresponds to a physical process such as strain hardening and annealing. Therefore, let us designate a material as perfect if

$$\dot{\underset{\sim}{V}}(\tau) = 0 \qquad (3.13)$$

for all $\tau \forall\ 0,t$. Alternatively, equation (3.13) implies that the history of $\underset{\sim}{V}(t)$ must vanish, i.e.

$$\underset{\sim}{V}^t(s) = 0 \ \forall \ t > 0 \qquad (3.14)$$

and from equation (2.13) it then follows that

$$\underset{s=0}{\overset{\infty}{\underset{\sim}{Q}}} [\underset{\sim}{R}^t(s) + {*}\underset{\sim}{F}_d^t(s)] = \underset{s=0}{\overset{\infty}{\underset{\sim}{Q}}} [\underset{\sim}{R}^t(s)] \qquad (3.15)$$

for all t>0.

To investigate the consequence of (3.15) let us introduce the difference histories of the deformation defined as

$$\begin{aligned} {*}\underset{\sim}{F}^t(s) &= \underset{\sim}{F}^t(s) - \underset{\sim}{F}(t) \\ {*}\underset{\sim}{R}^t(s) &= \underset{\sim}{R}^t(s) - \underset{\sim}{R}(t) \end{aligned} \qquad (3.16)$$

recalling

$${*}F_d^t(s) = \underset{\sim}{F}_d^t(s) - \underset{\sim}{F}_d(t)$$

with the property ${*}\underset{\sim}{F}^t(0)={*}\underset{\sim}{R}^t(0)={*}\underset{\sim}{F}_d^t(0)=\underset{\sim}{0}$. The constitutive functional $\underset{\sim}{Q}$ can be redefined so that the stress is given by

$$\underset{\sim}{\delta}(t) = \underset{s=0}{\overset{\infty}{\underset{\sim}{Q}}} [{*}\underset{\sim}{F}^t(s);\underset{\sim}{F}(t)] \qquad (3.17)$$

since the knowledge of the history ${*}F^t(s)$ and $\underset{\sim}{F}(t)$ is equivalent to the entire deformation history $\underset{\sim}{F}^t(s)$. Thus, equation (3.15) becomes

$$\underset{s=0}{\overset{\infty}{\underset{\sim}{Q}}} [{*}\underset{\sim}{R}^t(s) + {*}\underset{\sim}{F}_d^t(s);\underset{\sim}{R}(t)] = \underset{s=0}{\overset{\infty}{\underset{\sim}{Q}}} [{*}\underset{\sim}{R}_d^t(s);\underset{\sim}{R}(t)] \qquad (3.18)$$

for all t>0. Since equation (3.18) must hold for all choices of ${*}F_d^t(s)$ the material functional is history independent.

The quantity $\underset{\sim}{Q}$ then must be a function of the current deformation rather than a functional of the deformation history. Thus for materials where equation (3.14) is satisfied the constitutive equation (2.8) becomes

$$\underset{\sim}{S}(t) = \underset{\sim}{Q}[\underset{\sim}{F}(t)]. \tag{3.19}$$

Equation (3.19) the description of an elastic solid. Note that this development does not imply that elastic solids are free from damage; but, rather a constitutive equation of the form of equation (3.19) can not account for material degradation.

d) Relationship to Dissipation

A further property of the material variation tensor should now be presented; namely, the relationship between $\underset{\sim}{V}$ and dissipation. The dissipation integral defined by Coleman [4] is

$$I(t_1,t_o) = \int_{t_o}^{t_1}[\frac{1}{p}\,\mathrm{tr}(\underset{\sim}{t}\cdot\underset{\sim}{L}) - \eta\,\dot{\theta}]dt \tag{3.20}$$

for $t_1 > t_o$ where η and θ are the entropy and temperature, respectively. The dissipation integral must satisfy the integrated form of the dissipation inequality.

$$I(t_o,t_1) \geqq \Psi(t_1) - \Psi(t_o) \tag{3.21}$$

where Ψ is the Helmholtz free energy.

The stress power term can be rewritten using equation $(2.4)_2$ and (2.7) as

$$\frac{1}{p}\,\mathrm{tr}(\underset{\sim}{T}\cdot\underset{\sim}{L}) = \frac{1}{p}\,\mathrm{tr}(\underset{\sim}{L}\cdot\underset{\sim}{T}) = \frac{1}{p}\,\mathrm{tr}(\dot{\underset{\sim}{F}}\cdot\underset{\sim}{F}^{-1}\cdot p\underset{\sim}{F}\cdot\underset{\sim}{S}) = \mathrm{tr}(\dot{\underset{\sim}{F}}\cdot\underset{\sim}{S}). \tag{3.22}$$

Substituting (3.22) into (3.20) gives

$$I(t_1,t_o) = \int_{t_o}^{t_1} \mathrm{tr}(\dot{\underset{\sim}{F}}\cdot\underset{\sim}{S})dt \tag{3.23}$$

for an isothermal process. Equation (3.23) is now consistent with the previous assumption of a simple material.

Consider a deformation controlled cyclic process on the time interval [0,a]. Define $\underset{\sim}{S}_1(t)$ for $t\epsilon[0,a]$ to be stress response during the first cycle and let $\underset{\sim}{S}_2(t)$ for $t\epsilon[a,2a]$ represent the stress response during the second cycle such that

$$I_1 = \int_o^a \mathrm{tr}\ (\dot{\underset{\sim}{F}}\cdot\underset{\sim}{S}_1)dt$$

and

$$I_2 = \int_a^{2a} \mathrm{tr}\ (\dot{\underset{\sim}{F}}\cdot\underset{\sim}{S}_2)dt. \tag{3.24}$$

The change in the dissipation between the first and second intervals is then

$$\Delta I_{21} = {}_{2} - {}_{1} = \int_{o}^{a} \mathrm{tr}\ \dot{\underset{\sim}{F}}\cdot(\underset{\sim}{S}_2-\underset{\sim}{S}_1)\ dt; \qquad (3.25)$$

however, from equation (2.12) it follows that

$$\Delta I_{21} = \int_{o}^{a} \mathrm{tr}(\dot{\underset{\sim}{F}}\cdot\underset{\sim}{V}_{21})dt. \qquad (3.26)$$

where the material variation tensor $\underset{\sim}{V}_{21}=\underset{\sim}{S}_2-\underset{\sim}{S}_1$. Equation (3.26) shows that the change in dissipation is linear in the material variation tensor $\underset{\sim}{V}$ for an isothermal cyclic process.

This result has a simple but useful interpretation for a one-dimensional deformation. Consider two consecutive isothermal hysteresis loops as shown in Figure 2. The dissipation change from (3.26) can be written as

$$\Delta I_{21} = \int_{o}^{a} V_{21}\ dF = \int_{o}^{a} (S_2-S_1)dF, \qquad (3.27)$$

where ΔI_{21} is the area between the curves. Observe that if the dissipation change $\Delta I_{21}=0$ then the two loops are identical and by definition the material variation $\underset{\sim}{V}$ is zero. Consequently this shows that if $\underset{\sim}{V}$ vanishes the dissipation does not necessarily vanish. However, since $\underset{\sim}{V}$ does vanish we know that the constitutive relationship is history independent; thus for a material with dissipation equation (3.19) must be replaced by a rate type constitutive relationship of the form

$$\underset{\sim}{S} = \underset{\sim}{P}[\underset{\sim}{F}, \underset{\sim}{F}^{(1)}, \underset{\sim}{F}^{(2)}, \ldots \underset{\sim}{F}^{(m)}] \qquad (3.28)$$

where $\underset{\sim}{F}^{(m)}$, m=0, 1, 2..., are m material time derivatives of the deformation gradient $\underset{\sim}{F}$. Equation (3.28) is not a simple material as originally assumed; however, it is consistent with the theories of rubber elasticity, Rivlen [7]. Of course if I=0, then $\underset{\sim}{V}=0$ and equation (3.16) is acceptable as a constitutive model.

e) Application to Experimental Procedures

The constitutive equation for a simple material, i.e.

$$\underset{\sim}{S} = \frac{1}{p}\underset{\sim}{F}^{-1}\cdot\underset{\sim}{T} = \underset{s=0}{\overset{\infty}{\underset{\sim}{Q}}}[\underset{\sim}{F}^{t}(s)], \qquad (3.29)$$

must in general be determined from an experimental program. This is usually done by assuming an appropriate representation for $\underset{\sim}{Q}$, consistent with the class of materials being studied, and allowing a number of parameters in this representation to be adjustable. A set of experiments are then conducted to determine the response $\underset{\sim}{T}$ to the controlled deformation $\underset{\sim}{F}$. Substituting the experimental data into (3.29) gives a set of equations to solve for the experimental parameters in the assumed constitutive model. This procedure, or variations of it, are standard practice by experimentalist; however, the method becomes difficult, if not impossible, as the complexity of the material increases. In particular, for history dependent materials this process becomes very difficult because the experimental procedure matches a single response to each controlled deformation history and provides no rule for the effect of change deformation histories.

Consider now the material variation tensor $\underset{\sim}{V}$ as calculated from equation (3.5) and (3.7); i.e.

$$\underset{\sim}{V}(t) = \int_0^t \{\delta \underset{\sim}{\overset{\infty}{Q}}_{s=0} [F^t_D(s) | \dot{\underset{\sim}{F}}{}^t_D(s)] - \delta\underset{\sim}{\overset{\infty}{Q}}_{s=0} [\underset{\sim}{R}^t(s) | \dot{\underset{\sim}{R}}{}^t(s)]\}dt. \qquad (3.30)$$

The material variation tensor $\underset{\sim}{V}$ can also be determined experimentally from a controlled set of deformation histories $\underset{\sim}{F}{}^t_D(s)$ and $\underset{\sim}{R}^t(s)$. Substitution of this data back into (5.1) will provide information to determine an additional set of parameters in the constitutive model. However, in this case the governing equation involves the Fréchet derivative $\delta\underset{\sim}{Q}$, which is the change in the constitutive functional due to the difference history $\underset{\sim}{F}{}^t_d(s) = \underset{\sim}{F}{}^t_D(s) - \underset{\sim}{R}^t(s)$. Thus a direct measure of the history effects are included in the experimental program. This should provide a method to increase the accuracy of the constitutive model as well as investigating history effects such as fading memory.

Generalization to Other Constitutive Relationships

In the previous two sections the material variation tensor V(t) was introduced as characterizing the change of state of a simple material due to an arbitrary deformation. Further it was left up to the reader to identify the change of state as a degradation or enhancement of the material properties. In this section these ideas are restated as a general concept applicable to other constitutive equations that are a functional of any number of histories.

In addition the material variation concept is restated to allow for increased sensitivity of the damage measure. Recall the result for cyclic loading. In this case the damage is observed by changes in the hysteresis loops. For some materials, i.e., metals, there are significant changes in the loop during the first few cycles but, often the loop approaches a stable shape long before failure. This indicates that the observed parameter, namely stress, is not sensitive to the changes in the material state for predicting catastrophic failure. However, other detection systems may be useful. Consider, for example, a system using dislocation density as failure criteria. In general the dislocation density is a functional of the deformation history and could be measured by an ultrasonic or optical probe. In this arrangement the observed parameter is a function of the value of the constitutive functional.

a) Generalization to Multiple Histories

Let $\underset{\sim}{N}(\underset{\sim}{x},t)$, $\underset{\sim}{A}_1(\underset{\sim}{x},t)$, $\underset{\sim}{A}_2(\underset{\sim}{x},t) \ldots \underset{\sim}{A}_N(\underset{\sim}{x},t)$ represent scalar, vector or tensor functions defined on the spatial coordinates $\underset{\sim}{x}$ in B and $t\varepsilon(-\infty,\infty)$. Assume the response $\underset{\sim}{N}$ is determined by a functional $\underset{\sim}{P}$ of the histories of the functions $\underset{\sim}{A}_1 \ldots \underset{\sim}{A}_N$; that is

$$\underset{\sim}{N}(t) = \mathop{\underset{\sim}{P}}\limits_{s=0}^{\infty} [\underset{\sim}{A}_1^t(s),\ \underset{\sim}{A}_2^t(s),\ldots \underset{\sim}{A}_N^t(s)]. \qquad (4.1)$$

Equation (4.1) may represent a constitutive equation characterizing any physical, chemical, electrical, thermal, or mechanical property of a system. The $\underset{\sim}{A}_1 \ldots \underset{\sim}{A}_N$ are any set of relevant parameters. The functional $\underset{\sim}{P}$ can be redefined on the space of the vector

$$\underset{\sim}{\Gamma}(t) = [\underset{\sim}{A}_1(t),\ \underset{\sim}{A}_2(t),\ldots \underset{\sim}{A}_N(t)]. \qquad (4.2)$$

Then the constitutive equation (4.1) can be rewritten as

$$\underset{\sim}{N}(t) = \overset{\infty}{\underset{\sim}{P}}\ [\underset{\sim}{\Gamma}^t(s)]. \qquad (4.3)$$

Retracing the steps of Section II, let $\underset{\sim}{\Gamma}_R^t(s)$ represent a reference history with the properties

$$\underset{\sim}{\Gamma}_R^t(s) \begin{cases} = \underset{\sim}{0}\ \forall\ S\varepsilon(t,\infty) \\ \text{is arbitrary}\ \forall\ S\varepsilon\ 0,t\ . \end{cases} \qquad (4.4)$$

Next denote a deviation history $\underset{\sim}{\Gamma}_d^t(s)$ and require that

$$\underset{\sim}{\Gamma}_d^t(s) \text{ is arbitrary V } S\varepsilon(t,\infty)$$

and

$$\mathop{\underset{\sim}{P}}\limits_{s=0}^{\infty}\ [\underset{\sim}{\Gamma}_d^t(s)] = 0,\ \forall\ t\varepsilon\ 0,t\ . \qquad (4.5)$$

The total history $\underset{\sim}{\Lambda}^t(s)$ can now be defined as the composition

$$\underset{\sim}{\Lambda}^t(s) = \underset{\sim}{\Gamma}_R^t(s) + \underset{\sim}{\Gamma}_d^t(s) - \underset{\sim}{\Gamma}_d(t) = \underset{\sim}{\Gamma}_R^t(s) + {}^*\underset{\sim}{\Gamma}_d^t(s) \qquad (4.6)$$

for $s\varepsilon(0,\infty)$. The material variation tensor $\underset{\sim}{V}(t)$ indicates the change in the material property $\underset{\sim}{P}$ due to the deviation ${}^*\underset{\sim}{\Gamma}_d^t(s)$ from the reference history $\underset{\sim}{\Gamma}_R^t(s)$; that is,

$$\underset{\sim}{V}(t) = \mathop{\underset{\sim}{P}}\limits_{s=0}^{\infty} [\underset{\sim}{\Lambda}^t(s)\ \ -\ \mathop{\underset{\sim}{P}}\limits_{s=0}^{\infty}\ \underset{\sim}{\Gamma}_R^t(s)]. \qquad (4.7)$$

Alternately the material variation tensor can be determined from the derivative of the functional $\underset{\sim}{P}$ using equation (3.8) and (3.9)

$$\underset{\sim}{V}(t) = \int_o^t \dot{\underset{\sim}{V}}(\tau)d\tau = \int_o^t\{\mathop{\delta\underset{\sim}{P}}\limits_{s=0}^{\infty}[\underset{\sim}{\Lambda}^\tau(s)|\underset{\sim}{\Lambda}^\tau(s)]-\delta\overset{\infty}{\underset{\sim}{P}}\ [\underset{\sim}{\Gamma}_R^\tau(s)|\dot{\underset{\sim}{\Gamma}}_R(\tau)]\}d\tau \qquad (4.8)$$

The linearity properties of $\delta\underset{\sim}{P}$ permits the separation of variation of the material properties due to each of the histories $\underset{\sim}{A}_1,\ \underset{\sim}{A}_2 \ldots \underset{\sim}{A}_N$. Substituting (4.2) into (4.8) gives

$$\underset{\sim}{V}(t) = V_{\underset{\sim}{A}_1}(t) + V_{\underset{\sim}{A}_2}(t) + \ldots + V_{\underset{\sim}{A}_N}(t) \qquad (4.10)$$

where

$$V_{A_i} = \int_0^t \{\delta \overset{\infty}{\underset{s=0}{\underset{\sim}{P}}}[\underset{\sim}{\Lambda}^\tau(s)|\dot{\underset{\sim}{A}}{}^\tau_{id}(s)] - \delta \overset{\infty}{\underset{s=0}{\underset{\sim}{P}}}[\Gamma^\tau_{iR}(s)|\dot{A}^\tau_{iR}(s)]\}d\tau \qquad (4.11)$$

and

$$\underset{\sim}{\Gamma}_R = \underset{\sim}{A}_{1R}, \underset{\sim}{A}_{2R}, \ldots, \underset{\sim}{A}_{NR}$$

$$\underset{\sim}{\Gamma}_d = \underset{\sim}{A}_{1d}, \underset{\sim}{A}_{2d}, \ldots, \underset{\sim}{A}_{Nd} \,. \qquad (4.12)$$

Equation (4.11) shows that the damage or enhancement in the material property due to a particular history $A_i^t(s)$ will vanish if

$$\dot{\underset{\sim}{A}}{}^\tau_{id}(s) = \dot{\underset{\sim}{A}}{}^\tau_{iR}(s) = \underset{\sim}{0} \qquad (4.13)$$

or if the parameter $\underset{\sim}{A}_i$ is constant for throughout both experiments. Further notice the coupling that exists between the N histories $\underset{\sim}{A}_i$. The first argument of $\delta\underset{\sim}{P}$ is the total vector $\underset{\sim}{\Lambda}$ or $\underset{\sim}{\Gamma}$ where the second argument is only one component of Λ or Γ. This shows that if Equation (4.13) is satisfied for the component $\underset{\sim}{A}_i$, the constant value of $\underset{\sim}{A}_i$ is still coupled to the remaining variation components in Equation (4.10).

The generalization of the concepts material variation to an arbitrary functional dependence is seen to follow directly once the form of coupling between terms is understood; therefore, for convenience, the remainder of this manuscript limited to a constituting relationship with a single vector history as given in Equation (4.3).

b) Alternative Sensing Parameters

Let $\underset{\sim}{N}(t)$, as defined by Equation (4.3), be some material response characteristic. Further, assume $\underset{\sim}{N}(t)$ is observed by observing changes in the parameter $\underset{\sim}{M}$ through the constitutive relationship

$$\underset{\sim}{M}(\underset{\sim}{x},t) = f[\underset{\sim}{N}(\underset{\sim}{x},t)] \qquad (4.14)$$

that exist for all x in B and $t\varepsilon(-\infty,\infty)$. For example, $\underset{\sim}{M}$ could be the intensity of a reflected ultrasonic wave form and $\underset{\sim}{N}$ could be a dislocation density as discussed earlier. Thus, the observable response characteristic is related to the generalized deformation history $\underset{\sim}{\Gamma}^t(s)$ by

$$\underset{\sim}{M}(\underset{\sim}{x},t) = \underset{\sim}{f}\{\overset{\infty}{\underset{s=0}{\underset{\sim}{P}}}[\underset{\sim}{\Gamma}^t(s)]\}. \qquad (4.15)$$

The variation $\underset{\sim}{V}_M$ in the observed response parameter $\underset{\sim}{M}$ due to the difference in the deviation history $\underset{\sim}{\Lambda}_D$, is defined by

$$\underset{\sim}{V}_{\underset{\sim}{M}} = \underset{\sim}{f}\{ \mathop{\underset{\sim}{P}}\limits_{s=0}^{\infty} [\underset{\sim}{\Lambda}^{t}(s)]\} - \underset{\sim}{f}\{[\mathop{\underset{\sim}{P}}\limits_{s=0}^{\infty} \ \Gamma_{\underset{\sim}{R}}^{t}(s)]\}. \qquad (4.16)$$

However, using equation (3.8) the above can be rewritten as

$$\underset{\sim}{V}_{\underset{\sim}{M}} = \int_{o}^{t} \dot{V}_{\underset{\sim}{M}}(\tau)d\tau$$

$$= \int_{o}^{t} (\nabla_{\underset{\sim}{N}}\cdot \underset{\sim}{f})\cdot\frac{d}{d\tau}\{[\mathop{\underset{\sim}{P}}\limits_{s=0}^{\infty} \ \underset{\sim}{\Lambda}^{\tau}(s)] - \mathop{\underset{\sim}{P}}\limits_{s=0}^{\infty} [\Gamma_{\underset{\sim}{R}}^{\tau}(s)]\}d\tau$$

$$= \int_{o}^{t}(\nabla_{\underset{\sim}{N}}\cdot \underset{\sim}{f})\cdot\{\delta\mathop{\underset{\sim}{P}}\limits_{s=0}^{\infty} [\underset{\sim}{\Lambda}^{\tau}(s)|\dot{\underset{\sim}{\Lambda}}^{\tau}(s)] - \delta\mathop{\underset{\sim}{P}}\limits_{s=0}^{\infty} [\Gamma_{\underset{\sim}{R}}^{\tau}(s)|\Gamma_{\underset{\sim}{R}}^{\tau}(s)]\}d\tau$$

$$= \int_{o}^{t}(\nabla_{\underset{\sim}{N}}\cdot \underset{\sim}{f}) \cdot \dot{\underset{\sim}{V}}(\tau)d\tau \qquad (4.16)$$

where $\dot{\underset{\sim}{V}}$ is the variation rate of $\underset{\sim}{N}$ as defined in equation (4.8) and $\nabla_{\underset{\sim}{N}}\cdot \underset{\sim}{f}$ is a generalized gradient operation. Thus it is seen that $V_{\underset{\sim}{M}}$ is linear in $\underset{\sim}{V}$ and the scale amplification factor $\nabla_{\underset{\sim}{N}}\cdot \underset{\sim}{f}$.

ACKNOWLEDGEMENT

The author thanks Drs. T. Nicholas and J. Gallager of Wright Patterson Air Force Base and Professors A. Strauss, R. Dunholter, B. Berstein, and W. Krauss for the comments during the development of this manuscript. This work was partially supported by the Metals Behavior Branch, Metals and Ceramics Division, Air Force Materials Laboratory.

REFERENCES

1. Conway, J.B., Stentz, R.H. and Berling, J.T., Fatigue, Tensile, and Relaxation Behavior of Stainless Steel. TID 26135, National Technical Information Service, U.S. Dept. of Commerce, Springfield, Virginia (1975).

2. Berkovits, A., "Holographic Approach to Predicting Inelastic Strain at High Temperature". DASA Technical Note, NASA-TN-D6937, September 1972.

3. Truesdell C. and Nall, W., "Non-Linear Field Theories of Mechanics", Hanbuch Du Physik, Band III/3, Ed. W. Flügge, Springer Verlag (1965).

4. Coleman, B.D., "Thermodynamics of Materials with Memory", Archive Rational Mechanics and Analysis, 17.1, (1964).

5. Volterra, Vito, Theory of Functionals and of Integral and Integro-Differential Equations, Blackie & Son Limited, London, (1930).

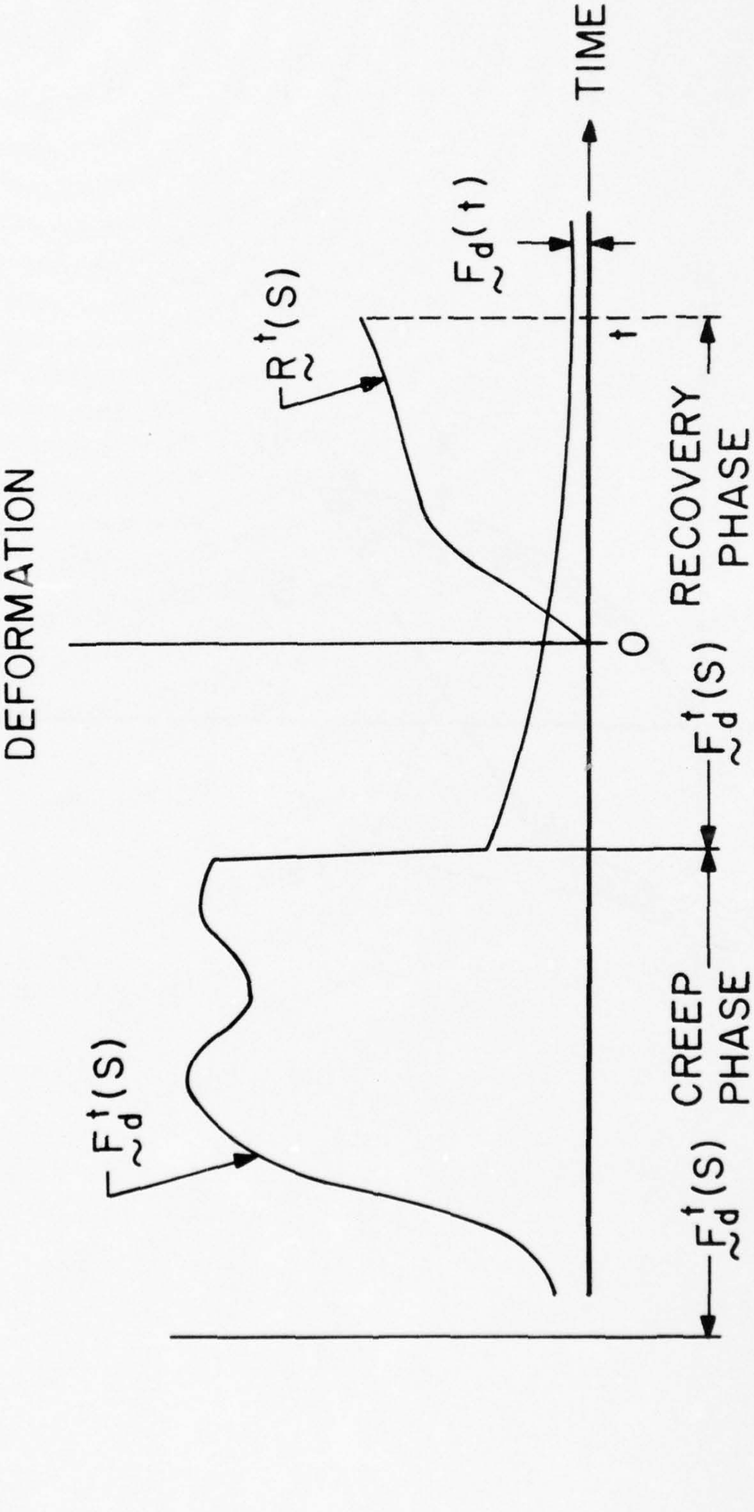
6. Reisz, F. and Nagy, B.Sz., Functional Analysis, Trans. by L.F. Boron, Rederick Ungar, New York, (1965).

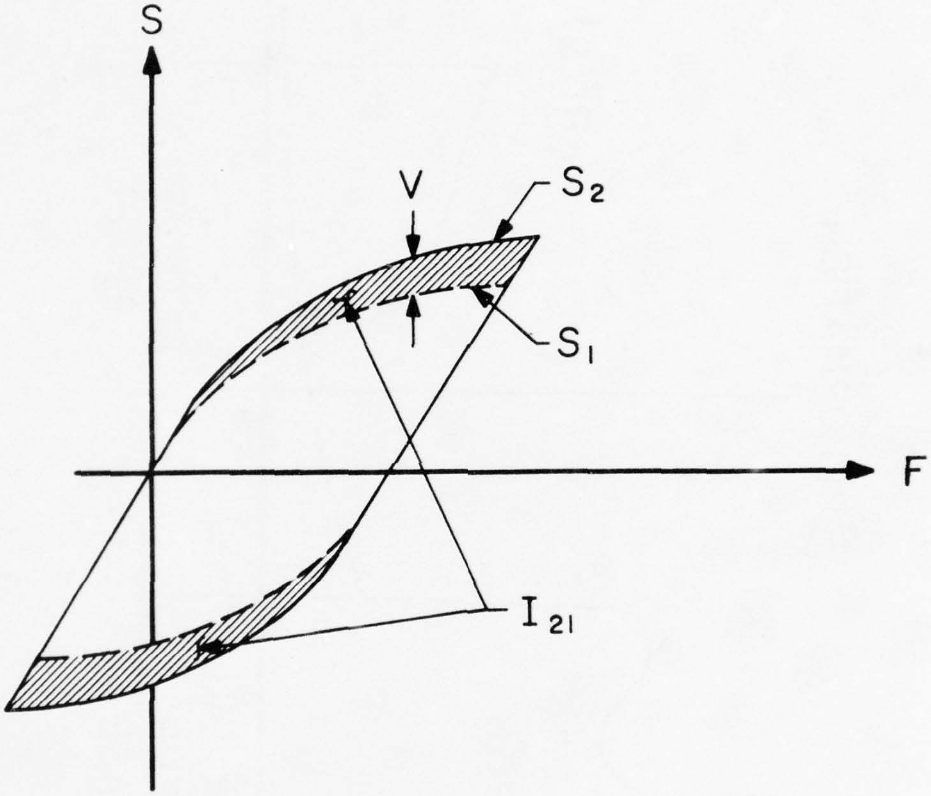
7. Rivlin, R.S. and Erickson, J.L., "Stress-Deformation Relations for Isotropic Materials", Journal of Rational Mechanics and Analysis, 4.2, (1955).

FIGURES

Figure 1. Definition of the reference deformation history $\underset{\sim}{R}^t(s)$ and the preworking deformation history $\underset{\sim}{F}^t_d(s)$.

Figure 2. Two successive hysteresis loops showing the change is dissipation I_{21}.


DEFORMATION
TIME
$\underset{\sim}{F}_d(t)$
$\underset{\sim}{R}^t(S)$
t
O
$\underset{\sim}{F}_d^t(S)$
$\underset{\sim}{F}_d^t(S)$
$\underset{\sim}{F}_d^t(S)$
CREEP PHASE
RECOVERY PHASE


S
V
S2
S1
F
I 21

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

CONTINUOUS PERFORMANCE MEASUREMENT OF MANUALLY CONTROLLED SYSTEMS

Prepared by:	Richard A. Miller Phd.
Academic Rank:	Assistant Professor
Department and University:	Department of Industrial and Systems Engineering Ohio State University
Assignment:	
(Laboratory)	Aerospace Medical Research Lab
(Divisions)	Human Engineering and Environmental Medicine
(Branches)	Systems Evaluation and Systems Technology
USAF Research Colleague:	Jerry P. Chubb Carroll N. Day
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

CONTINUOUS PERFORMANCE MEASUREMENT

of

MANUALLY CONTROLLED SYSTEMS

By

Richard A. Miller

ABSTRACT

A method of computing the performance of a human controller in a man-machine control system is described, evaluated and critiqued. The performance measurement technique is based on optimal control theory and is essentially a comparison of the control used by the operator at any point in time and state with the optimal control that would be applied at the same time-state point. To be useful the performance information must be available continuously in real time.

The continuous performance measurement concept is briefly reviewed and computational requirements obtained. It is shown that the method is applicable, given the state of the art in control theory and current computational capabilities, only for simple linear systems with quadratic performance measures. This restricts its utility to simple compensatory laboratory tracking tasks.

Problems of continuous performance measurement in realistic systems in realistic task environments are also briefly discussed.

INTRODUCTION

Control theoretic methods have long been used to model and analyze the response of a human operator performing a manual control task. Classical control models are well established and optimal control models are becoming more common (1,2). These models provide convenient ways of parameterizing the response of the human operator to a subset of the various stimuli to which he is subjected. The models are descriptive in the sense that they mathematically represent what an operator did in a particular task.

Recent work (3) has attempted to use optimal control methods in a novel way to provide real-time continuous performance feedback to the operator concerning how well he is performing the control task with respect to some specified measure of effectiveness. Traditionally, performance feedback to the operator is in the form of summary statistics computed after the task is completed. In many situations, such as experiments to measure excess operator capacity with secondary tasks, it is desireable to have continuous measurements of the level of performance on the primary task. Such information can be used to adapt the level of difficulty of the secondary task in real-time. Continuous performance feedback might also be used to provide aiding to the operator.

The purpose of this work is to carefully review the work(3) in particular and the continuous performance measure problem in general to determine the class of manual control tasks for which it is applicable. Particular attention is paid to computational requirements for real-time implimentation and limits on system complexity.

The performance measurement problem is first reviewed in general terms. The optimal control methodolgy used is then reviewed and computational problems defined. Limitations of the current methodology are then reviewed and general, philosophical reservations are discussed.

CONTINUOUS PERFORMANCE MEASUREMENT

In the typical manual control system, a human operator manipulates controls to guide the response of some dynamic system (a vehicle). Mathematically the vehicle is modeled by a set of ordered time functions representing the possible vehicle responses and control manipulations for the time period of interest. Specifically, the control responses are modeled by a set of real valued m dimensional time function, i.e.

$$U:T \rightarrow R^m \qquad (1)$$

U is the set of control responses, R denotes the real numbers, and T denotes the mathematical model of time. Here continuous time systems are considered and T is therefore assumed to be a subset of the real numbers

$$T = \left\{t : t_o \leq t \leq t_f ; \; t_o, t_f \in R\right\}$$

The index m is the number of controls available to the operator.

The vehicle is represented by the relation

$$V \subset U \times X \qquad (2)$$

where

$$X : T \rightarrow R^n \qquad (3)$$

T is again time and X denotes the set of vehicle state trajectories. The index n is the number of state variables used to describe the vehicle. Intuitively, the vehicle model V is a set of set control responses paired with vehicle state responses. The vehicle model is generally constructively specified with a set of differential equations, i.e.

$$V = \left\{(x,\upsilon) \in X \times U \, : \, \dot{x}(t) = f(x(t),u(t),t), \; t \in T\right\} \qquad (4)$$

With an optimal control model of the human operator, one assumes that the operator selects controls so as to maximize (or minimize) a measure of system performance subject to system constraints and available information. Abstractly a *performance measure* is any function which associates a number with each appearance of the system model. That is,

$$J : V \rightarrow R \qquad (5)$$

For systems of the type defined in equation (4), J is typically defined as follows:

$$J(x,u) = \int_{t_o}^{t_f} E(x(t),u(t),t)dt \qquad (6)$$

The optimal control problem solution, if one exists, is a function of the vehicle initial state. The optimal control problem solution is therefore conveniently represented as a relation

$$U^* = \left\{(x(t_o),u^*) \in R^n \times U\right\} \qquad (7)$$

which associates an optimal control u^* (or controls) with the initial vehicle state $x(t_o)$. If the human operator actually does perform like an optimal controller he will generate the relation U^*. U^*therefore is the model of the operator in an optimal control formulation of the manual control problem.

It is very important to notice that the performance measure is defined over the entire time set T and correspondingly the optimal control is a time function over the time set T. That is, in using the optimal control framework, the analyst specifies a time interval of interest (the duration of the task) and the resulting optimal control is a function defined over that interval. This simple fact is the origin of a major difficulty with the continuous performance measure methodology.

The relation U^* is an open loop solution which assumes the vehicle model is perfect in the sense that all environmental disturbances are accounted for. This is obviously unrealistic. The operator himself will introduce some uncertainty into the system. The effect of disturbances or uncertainties not accounted for in the vehicle model (2) is that the system state at some time point $t_1 \in T$ will deviate from that anticipated from (2) and adjustments in the control applied should be made. To be consistent with the optimal control framework, these adjustments must be made in a manner which optimizes performance over the remaining time interval $t_1 \leq t \leq t_f$. Formally, this requires defining a control problem for each point in time-state space. This is precisely a dynamic programming formulation of the control problem.

The vehicle model applicable in the time interval T_t,

$$T_t = \{\tau : t \leq \tau \leq t_f\}$$

is the restriction of V to that time interval, i.e.

$$V_t \subset U_t \times X_t \qquad (8)$$

$$U_t : T_t \rightarrow R^m \qquad (9)$$

$$X_t : T_t \rightarrow R^n \qquad (10)$$

The performance measure is

$$J_t : V_t \rightarrow R \qquad (11)$$

$$J_t(x_t, u_t) = \int_t^{t_f} E(x(s), u(s), s)ds \qquad (12)$$

The solution relation then takes the form

$$U_t^* = \{(x(t), u_t^*) \in R^n \times U_t\} \qquad (13)$$

There is a relation U_t^* for each time point $t \in T$. The relation (13) defines the optimal control to apply over the interval $t \leq \tau \leq t_f$ given the vehicle state at time t is x(t).

The maximum (or minimum) value of J_t is dependent only on the state at the time t. The optimal control is given by equation (13) and can be expressed as

$$u_t^* = U_t^*(x(t)) \in U_t \qquad (14)$$

The corresponding optimal state trajectory is

$$x_t^* = V_t(u_t^*) = V_t(U_t^*(x(t)) \qquad (15)$$

Therefore, a function

$$G^*: T \times R^n \rightarrow R \qquad (16)$$

can be defined as follows:

$$\begin{aligned} G^*(t,x(t)) &= J_t(x_t^*, u_t^*) \\ &= J_t(V_t(U_t^*(x(t))), U_t^*(x(t))) \qquad (17) \end{aligned}$$

G^* is the optimal "cost to go" function and provides the optimal performance level possible from any point (t,x(t)) in time-state space $T \times R^n$.

The apparatus needed to develop the continuous measure of operator performance has now been developed. Assume that the operator applies a control $u \in U$ and the vehicle state follows a trajectory $x \in X$. At time t the system is in some x(t) and the performance level using this control for the interval $t \leq \tau \leq t_f$ is

$$J_t(x_t, u_t) = \int_t^{t_f} E(x(s), u(s), s)ds \qquad (18)$$

The optimal performance level is obtained from(17) and the degradation in performance is

$$J_t(x_t, u_t) = G^*(t, x, (t)) - J_t(x_t, u_t) \qquad (19)$$

Equation (19)is not realizable in real time since J_t requires knowledge of the control and state trajectory over the future time interval. The time derivative of J_t is however theoretically realizable.

$$\frac{d\ J_t(x_t,u_t)}{dt} = \frac{dG^*(t,x(t))}{dt} - \frac{dJ_t(x_t,u_t)}{dt} \qquad (20)$$

and using (18),

$$\frac{d\ J_t(x_t,u_t)}{dt} = \frac{dG^*}{dt}(t,x(t)) + E(x(t),u(t),t) \qquad (21)$$

Equation (21) requires knowledge only of the current system state and current control, assuming G^* is known. To emphasize this fact, we rewrite (21) as follows:

$$(x(t),u(t)) = \frac{dG^*}{dt}(t,x(t)) + E(x(t),u(t),t) \qquad (22)$$

Equation (22) has been suggested by Connelly and Zeskind (ref 3) as a continuous measure of the operator's performance.

The function has three properties that make it particularly attractive as a continuous performance measure. First its integral over T is precisely the performance difference between optimal control and that used by the operator. Second, (x(t),u(t)) is zero if u(t) is the optimal control corresponding to x(t). Third, it does not penalize current or future behavior for suboptimal control in the past. It tells the operator if he is doing the best he can given the current system state.

In summary, equation (22) provides the theoretical structure for a continuous measure of a human operator's control performance if the system he is controlling is a continuous time dynamical system and system performance objectives can be quantified by a mapping of the form displayed in equation (6). We now examine the practicality of this measure.

COMPUTATIONAL REQUIREMENTS

A quick review of the technical features of the previous section will indicate that given a vehicle model V and a system performance function J the major step in developing the continuous measure is the determination of the optimal cost to go function G^* (see equation 16). That is, the optimal performance level for <u>every</u> point in $T \times R^n$ must be known and it must have a derivative with respect to time. This means that G^* must either be a simple analytical form or that it be obtained by computational means. If computational methods are required the grid size used to discretize $T \times R^n$ becomes important and the dimension of the system state space (n) must be quite small (three or four maximum). Analytic forms are known only for linear systems and quadratic performance measures. That is V is a linear system defined by

$$\dot{x}(t) = A(t)x(t) + Bu(t) \qquad (23)$$

and

$$J = \int_{t}^{t_f} (x'(t)Qx(t) + u'(t)Ru(t))dt \qquad (24)$$

This is the same basic structure used in ref.(1,2) without stochastic components or limitations of human response incorporated.

Nonlinear systems and/or non-quadratic performance measures could not be calculated in real time and G would certainly have to be precomputed. The dimensionality and discretization issues become very important. It should be pointed out that the calculation involved is the solution of an optimal control for each point in $T \times R^n$, no small effort.

On a purely technical level, the continuous performance method is limited by our ability to solve optimal control problems. If optimal performance can be computed one can determine degradation from optimal.

MODELING ISSUES

The conceptual appeal of optimal control based continuous performance measures is that the operator is compared with a globally optimal standard. The problem is that if the global optimum can be defined and obtained, the human operator in the system is probably not necessary. That is, if the optimal response for every system state is known, why not impliment that action directly. The continuous performance measurement application of optimal control should be compared with other uses of optimal control in manual control modeling. It is typically assumed that the well trained operator will perform very close to optimal but that he adapts his response mechanism and hence <u>his</u> objective function to fit the task environment. The parameters in the performance measure are estimated from response data and are used to indicate how the operator adapted to the environment.

The key reason for including human operators in most systems is their adaptability and discrete decision making capability. The method suggested in ref. (3) does not lend itself to monitoring adaptation unless system performance measures appropriate for each condition to which the human must respond are determined beforehand. These conditions cannot generally be modelled as state changes in a continuous dynamic system such as that defined by equation (4). Such environmental changes more often involve changes in the structure of the controlled system (e.g. due to damage) or changes in the nature of the task performed by the operator.

Even restricting attention to simple control tasks, in the course of a mission a pilot is required to perform several tasks including climbs, cruise,rendevous,weapon delivery, etc. The control strategy for each is quite different, requiring a different performance measure. One overall mission performance measure based on control is not easily defined. This is particularly true when the technical constraints discussed in the previous section are considered.

The implication is that the method of continuous performance measurement considered is restricted to a subset of the tasks involved in mission segments. Overall performance with respect to mission requirements must be assessed with a more complex structure.

SUMMARY AND CONCLUSIONS

The mathematical structure of a type of continuous performance measure was presented and evaluated. It was determined that the method is limited by ones technical ability to solve optimal control problems. It is therefore restricted in practice to simple linear systems with quadratic performance measures. The method cannot at present be used with more realistic system performance measures of with complex nonlinear systems.

The work does however provide insight into the properties that measures of human operator performance chould have. Unfortunately the adaptability of the human and his discrete decision making capability cannot be measured within this structure.

Progress in computational methods,more comprehensive models of human performance in complex task environments must be developed, and theoretical advances in the control of complex systems are all necessary before continuous performance measures of this type will be realistic. Models developed to date provide precious little insight into the total integrated behavior of the human as a controller, information processor and decision maker.

REFERENCES

1. Kleinman,D.L., S. Baron, and W. H. Levison, "An Optimal Control Model of Human Response,Part 1: Theory and Validation," Automatica, Vol. 6, 1970

2. Phatak, A., H. Weinert, and I. Segall, Identification of the Optimal Control Model for the Human Operator, Final Report, F33615-73-C-4021, Systems Control Inc., Palo Alto, Calif., Sept 1974.

3. Connelly, E. M., and Zeskind, R. M., Development of New Techniques for the Measurement of Systems, Final Report, F33615-75-C-5088, Omnemii, Inc., Springfield, Va., June 1975.

4. Athans, M.,and P. Falb, Optimal Control, McGraw-Hill, 1966.

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

DIGITAL AUTOPILOT DESIGN

Part I: A Common Review of Frequency Domain Theory for Continuous and Discrete Time Linear Time Invariant Systems.

Part II: A Review of Methods of Generating Discrete Systems from Prototype Continuous Systems.

Prepared by:	James F. Delansky, PhD.
Academic Rank:	Associate Professor
Department and University:	Department of Electrical Engineering Penn State University
Assignment:	
(Laboratory)	Armament Laboratory
(Division)	Digited Guided Weapons
(Branch)	Systems Analysis and Simulation
USAF Research Colleague:	Major K. A. (Al) Gale
Date:	August 15, 1975
Contract No:	F44620-75-C-0031

DIGITAL AUTOPLIOT DESIGN

By

James F. Delansky

ABSTRACT

Approximately the first two weeks were used to obtain an overview of DLMA job tasks. The modular guided bomb projects were singled out. It was decided to concentrate on the use of a digital autopilot in these systems with the following as objectives of the research:

(1) To produce a relatively short document about digital filtering that would be useful to DLMA personnel in their future work with digital autopilots. It is felt that the best way to do this is to use the well known continuous system background of most personnel; therefore, this document will retrace continuous system theory and, at each step, show the corresponding step in discrete system theory.

(2) Look at various ways of specifying a digital autopilot, e.g.:

(a) Copy existing continuous analog autopilot. There are several ways of doing this.

(b) Specify the digital autopilot directly from the aerodynamic equations of the structure, possibly using ideas of state space and optimal control.

(3) Look at other ways of implementing the autopilot, e.g.:

(a) Synchronous RC filters.

(b) N-path filters.

Part I. A Common Review of Frequency Domain Theory for Continuous and Discrete Time Linear Time Invariant Systems.

INTRODUCTION

The purpose of this report is to provide, in a form that is as compact as possible, the necessary background to understand discrete - time linear time invariant systems and to investigate some design procedures for such systems.

Most technically trained people feel quite comfortable when talking about continuous - time linear time invariant (CT,L,TI) systems. A brief review of such a system, say, for simplicity, the single input - single output system shown in Figure 1, will get us started.

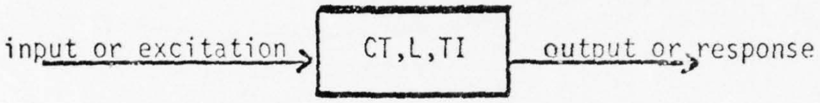


Figure 1

The continuous time means that the system operates on a continuous time input signal and produces a continuous time output signal. The linear means that the operation mentioned above is a linear one. The time invariant means that the operation does not change with time. Let *u(t) denote the class* of input signals, which for our purposes can be viewed as the map u: R → R, and let y(t) denote the class of output signals, y: R → R, where R denotes the real line. Now we have the condensed diagram of Figure 2.

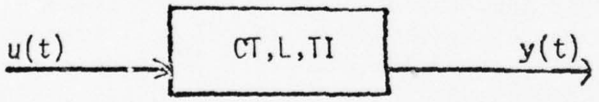


Figure 2

There are two basic problems in the theory of such systems.

1. The analysis problem: given the CT,L,TI system and the input u(t), determine the output y(t). Except for degenerate cases, which are not physically important, a solution exists and is unique.

2. The synthesis problem: given the input u(t) and the desired response y(t), find the CT,L,TI system. Here a solution may not exist, and even if it does, is generally not unique. The existence of a solution is of utmost importance; after all, you could work for a long time on a problem which had no solution! How can you tell when a solution exists? You must learn the capabilities of CT,L,TI systems. How? You master the analysis problem.

There are two ways to solve the analysis problem, (1) directly in the continuous time domain, or (2) indirectly via a frequency domain, as suggested by Figure 3.

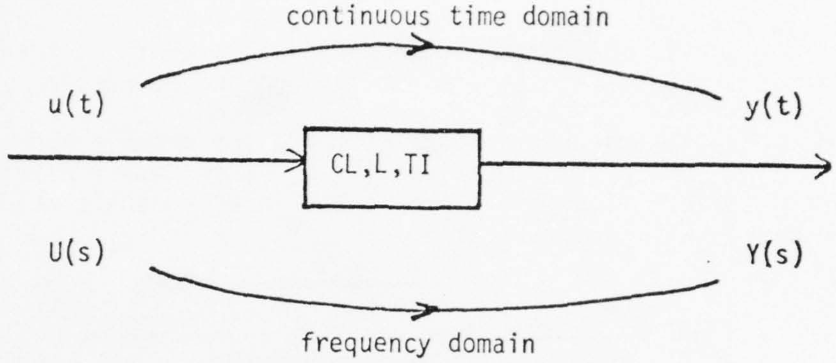


Figure 3

The continuous time domain approach could be viewed as solving the input-output relationship (for CT,L,TI systems, this is an ordinary linear constant coefficient differential equation) by well known classical methods, or by state equation formulation, or if only the forced response was desired, by a convolution integral. The frequency domain approach is essentially to find a transformation that will, say, trade the solution (y(t)) of a differential equation in the time domain for the solution (Y(s)) of an algebraic equation in the freauency domain. As suggested by the notation employed, the Laplace transform is the usual vehicle employed for CT,L,TI problems via the frequency domain.

Now we want to take the same approach toward discrete - time linear time invariant (DT,L,TI) systems as we did above for CT,L,TI systems. Let v(n) denote the class of input signals, v: $R_I \rightarrow R$, and let z(n) denote the class of output signals, z: $R_I \rightarrow R$, where R_I denotes the integer subset of the real line. So we have Figure 4, which is the discrete system counterpart to the continuous system Figure 3.

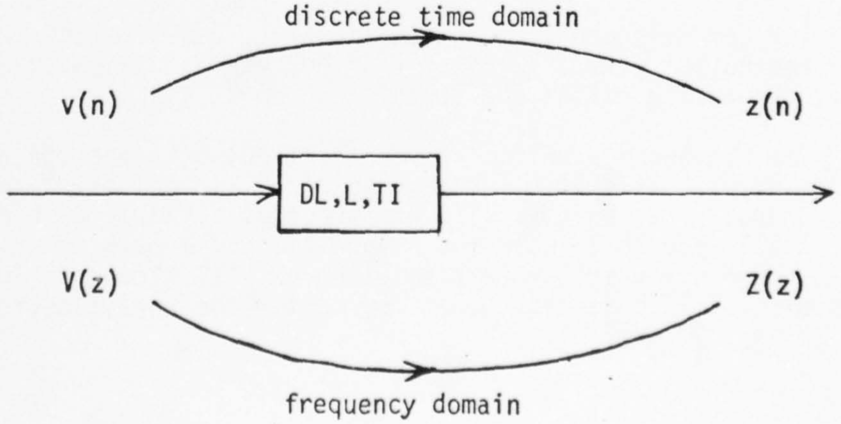


Figure 4

Now for each statement we made about CT,L,TI systems, a corresponding statement can be made about DT,L,TI systems.

As noted above, transform methods are an important means of study in CT,L,TI and DT,L,TI systems. Summaries of Fourier, Laplace, and Z - Transforms will be given in Sections 1, 2, 3, and 4 to provide a convenient reference in a form which relates these transforms within a common framework of notation and terminology. Relationships among these transforms are then given in Section 5. Theorems and properties will be stated without proof, since the references may be consulted when more completeness is desired.

The development of CT,L,TI frequency domain concepts usually starts with the theory of Fourier Series, then progresses to the Fourier Transform, and finally to the Laplace Transform. Similarly, the development DT,L,TI frequency domain concepts go from discrete Fourier Series, to discrete Fourier Transform, to the Z - Transform. Since the Fourier Series concepts are relatively uninteresting for our purposes, we will start with the Fourier Transform.

1. The Fourier Transform

A. Theorem 1-1. The conditions

1. f(t) real and piecewise continuous, $-\infty < t < \infty$

2. $\int_{-\infty}^{\infty} |f(t)|\, dt < \infty$, (i.e., f(t) is absolutely integrable)

3. f(t) bounded at discontinuities, imply

a. $$F(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\, dt \tag{1.1}$$

is defined for real ω and is continuous

b. $$I = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(j\omega)e^{j\omega t}\, d\omega \text{ (principle value)} \tag{1.2}$$

converges for all real t.

c. $$I = \begin{cases} f(t), \text{ where } f(t) \text{ is continuous} \\ \dfrac{f(t+) + f(t^-)}{2}, \text{ where } f(t) \text{ is discontinuous.} \end{cases}$$

B. The Fourier Integral (1.1) and the Inversion Integral (1.2) relate the function $F(j\omega)$ and f(t) uniquely under the hypothesis of Theorem 1-1. Thus the Fourier Transform is given by the pair

$$F(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\, dt = \mathcal{F}\left[f(t)\right] \tag{1.3}$$

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(j\omega)e^{jt\omega} d\omega = \mathcal{F}^{-1}\left[F(j\omega)\right] \tag{1.4}$$

C. Properties. (Assume indicated transforms exist.)

1. Linearity

$$\mathcal{F}\left[af(t) + bg(t)\right] = aF(j\omega) + bG(j\omega)$$

2. Differentiation of f(t)

Let $D \equiv \frac{d}{dt}$. If the $D^k f$, $k = 0, 1, 2, 3, \ldots$, are continuous and are absolutely integrable and $f(t) \rightarrow 0$ for $|t| \rightarrow \infty$,

Then

$$\mathcal{F}\left[D^k f(t)\right] = (j\omega)^k F(j\omega)$$

3. Integration of f(t)

If $D^{-k} f$, $k = 0, 1, 2, 3, \ldots$, and f(t) is piecewise continuous, then

$$\mathcal{F}\left[D^{-k} f(t)\right] = \frac{1}{(j\omega)^k} F(j\omega), \ \omega \neq 0.$$

4. Time shift.

$$\mathcal{F}\left[f(t-\tau)\right] = e^{-j\omega\tau} F(j\omega)$$

5. Frequency shift.

$$\mathcal{F}\left[e^{j\omega_0 t} f(t)\right] = F\left(j\left(\omega-\omega_0\right)\right)$$

6. Time multiplication.

$$\mathcal{F}\left[t^n f(t)\right] = j^n \frac{d^n F(j\omega)}{d\omega^n}$$

7. Time scale.

$$\mathcal{F}\left[f(at)\right] = \frac{1}{a} F\left(j \frac{\omega}{a}\right), \ a > 0.$$

D. Theorems

1. Parseval

If f(t) satisfies Theorem 1-1, then

$$\int_{-\infty}^{\infty} \left[f(t)\right]^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(j\omega)|^2 d\omega$$

2. Real convolution (complex multiplication)

If f(t) and g(t) satisfy Theorem 1-1 and are bounded, let

$$f * g = \int_{-\infty}^{\infty} f(\alpha)\, g\,(t-\alpha) d\alpha \text{ , then}$$

$$\mathcal{F}\left[f*g \right] = F(j\omega)\ G(j\omega)$$

3. Complex convolution (real multiplication)

For $F*G = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\beta)G(\omega-\beta)d\beta$, then

$$\mathcal{F}[fg] = F*G.$$

E. Brief tables of Fourier transforms,

Some transform pairs will be useful to show relationships among the various transforms. Two groups of functions are considered. Those in the first group (Table I.E.I) satisfy Theorem I-I and those in the second group (Table I.E.2) do not. The latter are properly treated as generalized functions. The singularity functions are denoted as follows (where $D^k \equiv \frac{d^k}{dt^k}$);

$\vdots$

$u_{-1}(t) = D\,u_0(t)$, unit doublet

$u_0\ (t) \equiv$ unit impulse

$u_1\ (t) = D^{-1}u_0(t)$, unit step

$\vdots$

Table 1.E.1

$f(t)$	$F(j\omega)$
1. $u_1(t+T)-u_1(t-T)$	$2T\,\dfrac{\sin\omega T}{\omega T}$
2. $\epsilon^{-at}u_1(t)$	$\dfrac{1}{a+j\omega}$
3. $\epsilon^{-a\|t\|}$	$\dfrac{2a}{a^2+\omega^2}$
4. $1-2(\|t\|/T), \quad \|t\| \leq T/2$ $0 \qquad\qquad , \quad \|t\| > T/2$	$\dfrac{T}{2}\left[\dfrac{\sin\frac{\omega T}{4}}{\frac{\omega T}{4}}\right]^2$
5. $2\omega_o\,\dfrac{\sin\omega_o t}{\omega_o t}$	$2\pi[u_1(\omega+\omega_o) - u_1(\omega-\omega_o)]$

Table t.E.2

f(t)	F(jω)
1. $u_o(t)$	1
2. $u_1(t)$	$\pi u_o(\omega) + \frac{1}{j\omega}$
3. K	$2\pi K u_o(\omega)$
4. $\cos \omega_o t$	$\pi[u_o(\omega-\omega_o) + u_o(\omega+\omega_o)]$
5. $\sin \omega_o t$	$-j\pi[u_o(\omega-\omega_o) - u_o(\omega+\omega_o)]$
6. $(\cos \omega_o t)u_1(t)$	$\pi/2[u_o(\omega-\omega_o) + u_o(\omega+\omega_o)] + \frac{j\omega}{\omega_o^2-\omega^2}$
7. $(\sin \omega_o t)u_1(t)$	$-j^{\pi/}2[u_o(\omega-\omega_o) - u_o(\omega+\omega_o)] + \frac{\omega_o}{\omega_o^2 - \omega^2}$
8. sgn t	$\frac{2}{j\omega}$
9. $\epsilon^{j\omega_o t}$	$2\pi u_o(\omega-\omega_o)$

2. The Laplace Transform

A. The Two-Sided (Bilateral) Laplace Transform is

$$F_2\,(s) = \mathcal{L}_2\left[f(t)\right] = \int_{-\infty}^{\infty} f(t)\; e^{-st}\, dt \qquad (2.1)$$

where $s = \sigma + j\omega$, if the integral converges for some σ such that $\sigma_1 < \sigma < \sigma_2$, where the real numbers σ_1 and σ_2 are dependent on f(t). Clearly this is an extension of the Fourier Integral (1.1) by inclusion in the integrand the "damping factor" $e^{\sigma t}$ so that the $e^{j\omega t}$ in the integrand of (1.1)replaced by $e^{\sigma t}\, e^{j\omega t} = e^{(\sigma+j\omega)\,t} = e^{st}$ in the integrand of (2.1)

The Inverse Laplace Transform is:

$$f(t) = \mathcal{L}_2^{-1}\left[F_2(s)\right] = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} F_2(s)\, e^{ts}\, ds \quad . \qquad (2.2)$$

where $\sigma_1 < \sigma < \sigma_2$.

The region of convergence of $F_2(s)$ must be considered to determine f(t). For the usual cases of signals that are bounded for large $|t|$, poles of $F_2(s)$ in LHP are associated with t>o and the poles of $F_2(s)$ in RHP are associated with t<o.

Theorem 2-1. The conditions
1. f(t) real, piecewise continuous, $-\infty < t < +\infty$
2. $\int_{-\infty}^{\infty} \left| f(t)\, e^{-\sigma t} \right| dt < \infty$, for some real σ

imply
a. $F_2(s) = \mathcal{L}_2\left[f(t)\right]$ exists
b. $f(t) = \mathcal{L}_2^{-1}\left[F(s)\right]$ where f(t) is continuous

A short table of Bilateral Laplace Transform pairs (for reasons given earlier) is given as Table 2.A.1.

Table 2 .A.1

f(t)	$F_2(s)$	
1. $\epsilon^{-a\|t\|}$, a > 0	$\frac{2a}{a^2-s^2}$,	$-a < \sigma < a$
2. $\frac{2a}{(t^2+a^2)}$, a > 0	$2\pi\epsilon^{-a\|s\|}$,	$\sigma = 0$
3. $u_1(t+T)-u_1(t-T)$ $= u_1(-t^2+T^2)$	$(2T)\,\frac{\sinh(Ts)}{(Ts)}$,	$\|\sigma\| < \infty$
4. $(2\omega_o)\,\frac{\sin(\omega_o t)}{(\omega_o t)}$	$2\pi u_1(s^2+\omega_o^2)$,	$\sigma = 0$

5. $\cos \omega_o t$ | $\pi u_o(s^2+\omega_o^2)$, $\sigma = 0$

Note the symmetry property illustrated by the pair 1. and 2. and the pair 3. and 4. Note that 5. does not satisfy the hypothesis of Theorem 2-1 it is simply another form of 4. of Table .E.2.

B. The One-Sided (Unilateral) Laplace Transform is used for functions $f(t)$ where $f(t) = o$ for $t < o$.
Emphasizing this by the notation $f(t)\, u_1(t)$, then

$$F(s) = \mathcal{L}\,[f(t)u_1(t)] = \int_0^\infty f(t)e^{-st}dt \tag{2.3}$$

where $s = \sigma + j\omega$, if the integral converges for some σ such that $\sigma > \sigma_o$.
The Inverse Laplace Transform is

$$f(t)\, u_1(t) = \mathcal{L}^{-1}\,[F(s)] = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} F(s)e^{ts}ds \tag{2.4}$$

Where $\alpha > \sigma o$. A unique transform pair is formed by (2.3) and (2.4).

Theorem 2-2. The conditions

1. $\int_0^T |f(t)|\,dt \leq K < \infty$, for some $T > 0$
2. $|f(t)| \leq Me^{\sigma_o t}$, for some $M > 0$ and $t > T$, implies

$F(s) = \int_0^\infty f(t)e^{-st}dt$ converges absolutely (and uniformly) for $\sigma > \sigma_o$

The interpretation of the lower limit in (2.3) is important when $f(t)$ or its derivatives are not continous at $t = o$. Then the lower limit must be specified as either 0 - or 0+. Either can be used provided it is used consistently.

Consider

$$\mathcal{L}_{o+}[\cos(\omega_o t)u_1(t)] = \frac{s}{s^2+\omega_o^2} \equiv F(s) \quad \text{and}$$

$$\mathcal{L}_{o-}[\cos(\omega_o t)u_1(t)] = \frac{s}{s^2+\omega_o^2} \quad , \text{then}$$

$$\mathcal{L}_{o+}\left[\frac{d}{dt}\left(\cos(\omega_o t)u_1(t)\right)\right] = \mathcal{L}_{o+}[-\omega_o \sin(\omega_o t)\, u_1(t) + \cos(\omega_o t)\, u_o(t)]$$

$$= \mathcal{L}_{o+}[-\omega_o \sin(\omega_o t)\, u_1(t)] + \mathcal{L}_{o+}[\cos(\omega_o t)u_o(t)]$$

$$= \frac{-\omega_o^2}{s^2+\omega_o^2} + 0$$

$$= s\,\frac{s}{s^2+\omega_o^2} - 1 = sF(s) - f(o+)$$

while $\mathcal{L}_{0^-}\left[\frac{d}{dt}\left(\cos(\omega_0 t)u_1(t)\right)\right] = \mathcal{L}_{0^-}\left[-\omega_0\sin(\omega_0 t)u_1(t) + \cos(\omega_0 t)u_0(t)\right]$

$$= \mathcal{L}_{0^-}\left[-\omega_0\sin(\omega_0 t)u_1(t)\right] + \mathcal{L}_{0^-}\left[\cos(\omega_0 t)u_0(t)\right]$$

$$= \frac{-\omega_0^2}{s^2+\omega_0^2} + 1 = \frac{s^2}{s^2+\omega_0^2} - 0$$

$$= sF(s) - f(0^-)$$

Theorem 2.3. The conditions

1. $F(s) = \int_0^\infty f(t)e^{-st}dt \quad , \quad \sigma > \sigma_0$

2. $I = \frac{1}{2\pi j}\int_{a-jb}^{a+jb} F(s)e^{ts}ds$

3. $a > \sigma_0$

4. f(t) of bounded variation imply

$$\lim_{b\to\infty} I = \begin{cases} 0, & t<0 \\ f\frac{(0+)}{2}, & t=0 \\ \frac{f(t+)+f(t-)}{2}, & t>0 \end{cases}$$

The Inversion Theorem 2-3 is the general method of computing inverse Laplace transforms, using the calculus of residues. Tables may also be used, because of the uniqueness of the Laplace transform pair. Of special usefulness in this regard is the partial fraction expansion technique.

Partial Fraction Technique

Let F(s) = N(s)/D(s), where N(s) and D(s) are polynomials such that m = degree of D > degree of N, then

1. Simple pole case

$F(s) = \sum_{j=1}^{n} \frac{A_j}{s-\lambda j}$, where the λj are simple zeros of D(s) and

$$A_j = \frac{N(s)}{D(s)}(s-\lambda_j)\Big|_{s=\lambda j}$$

2. Multiple pole case

$F(s) = \sum_{j=1}^{m}\sum_{i=1}^{mj} \frac{A_{ij}}{(s-\lambda_j)^i}$, where mj is

the order of the zero of D(s) at s = λj, and $n = \sum_{j=1}^{m} m_j$, and

$$A_{ij} = \frac{1}{(m_j-i)!}\frac{d^{m_j-i}}{ds^{m_j-i}}\left[\frac{N(s)(s-\lambda_j)^{m_j}}{D(s)}\right]\Big|_{s=\lambda j}$$

then the inverse transforms are, respectively,

1. $f(t) = \sum_{j=1}^{n} A_j e^{\lambda_j t} \quad , t>0$

2. $f(t) = \sum_{j=1}^{m} e^{\lambda_j t} \sum_{i=1}^{m_j} A_{ij} \frac{t^{i-1}}{(i-1)!} \quad , \quad t>0$

General method of evaluating integral I in Theorem 2-3: Jordan's Lemma and the Calculus of Residues.

Jordan's Lemma (A): The conditions

1. F(s) meromorphic for $\sigma \leq \sigma_o$

2. $|F(s)| \to o$ as $R\to\infty$ for $s = Re^{j\theta}$
   implies
   $I = \int_{C_\ell} F(s)\, e^{ts} ds \to o$ as $R \to \infty$, $t>o$

where C_ℓ is a semicircular path to the left with radius R and center at the origin of the s-plane.

Jordan's Lemma (B): The conditions

1. F(s) meromorphic for $\sigma \geq \sigma_o$

2. $|F(s)| \to o$ as $R \to \infty$ for $s = Re^{j\theta}$ implies

   $I = \int_{C_r} F(s) e^{ts} ds \to o$ as $R \to \infty$, $t<o$

where C_r is a semicircular path to the right with radius R and center at the origin of the s-plane.

Now for $\frac{1}{2\pi j} \int_C F(s)e^{ts}ds = \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} F(s)e^{ts}ds + \frac{1}{2\pi j}\int_{C_\ell} F(s)e^{ts}ds$

where C is the closed contour indicated on the right handside, Cauchy Residue Theorem implies that the left hand side is

$$\sum \text{residues of } F(s)\, e^{ts} \Big|_{\text{poles inside } C}$$

and Jordan's Lemma (A) implies the second term on the right (for suitable F(s)) is zero, then since $\sigma = \sigma_0$ is to the right of the poles of F(s),

$$\sum \text{residues of } F(s)e^{ts}\Big|_{\text{poles of } F(s)}$$

$$= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} F(s)e^{ts}ds \quad = f(t)\,,\ t>0$$

By similar agruement using Jordan's Lemma (B), f(t) = 0, t<0, since there are no poles to the right of $\sigma=\sigma_0$

Using similar arguements, the Inverse Two-Sided Laplace Transform of Section .A can be evaluated when the convergerce strip is given by $\sigma_1 < \sigma < \sigma_2$. For Jordan's Lemma (A) with $\sigma_0 = \sigma_1$.

$$\sum \text{residues of } F_2(s)\ e^{ts}\Big|_{\substack{\text{poles of } F_2(s)\\ \text{to the left of } \sigma_1}} = f(t),\ t>0$$

and using Jordan's Lemma (B) with $\sigma_0=\sigma_2$,

$$-\sum \text{residues of } F_2(s)\ e^{ts}\Big|_{\substack{\text{poles of } F_2(s)\\ \text{to the right of } \sigma_2}} = f(t),\ t<0$$

The partial fraction method can also be used with an analogous interpretation of the poles of $F_2(s)$ and associated components of f(t) for t>0 and f(t) t<0.

Finally the required residues are found from

1. Residue of F(s) $e^{ts} = \lim\limits_{s\to\lambda j} \{(s-\lambda j)\ F(s)\ e^{ts}\}$ for a simple pole of F(s) at $s = \lambda j$.

2. Residue of F(s) e^{ts}

$$= \frac{1}{(m-1)!} \lim_{s\to\lambda j}\left[\frac{d^{m-1}}{ds^{m-1}}(s-\lambda j)^m F(s)\,e^{ts}\right]$$

for an m - th order pole of F(s) at $s = \lambda j$.

Note that if the singularities of F(s) are not all poles, then a more general method of evaluating the inverse transform is required.

C. Properties of Unilateral transform.

1. Linearity

$$\mathcal{L}\{af(t) + bg(t)\} = a\,F(s) + bG(s)$$

2. Differentiation of F(t)

$\mathcal{L}\{Df(t)\} = F(s) - f(o)$,and

$\mathcal{L}\{D^k f(t)\} = s^k F(s) - s^{k-1} f(o) - \cdots - f^{(k-1)}\ (0)$

where $f^{(m)} = D^m f$.

4. Timeshift

$$\mathcal{L}[f(t-\tau)] = \begin{cases} e^{-\tau s}F(s) & , \tau > 0 \\ e^{\tau s}F(s) - e^{\tau s}\int_0^{\tau} f(t)\,e^{-st}\,dt & , \tau < 0 \end{cases}$$

5. Frequency shift

$\mathcal{L}[f(t)\ e^{at}] = F(s-a)$

6. Time multiplication

$\mathcal{L}[tf(t)] = -\frac{dF(s)}{ds}$

7. Time Scale

$\mathcal{L}[f(at)] = \frac{1}{a}\ F(s/a),\ a>0.$

D. Theorems

1. Integral Square

If f(t) satisfies Theorem 2-1 for $\sigma=0$, then

$$\int_{-\infty}^{\infty}[f(t)]^2\,dt = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} F_2(s)\,F_2(-s)\,ds$$

2. *Real Convolution*

$\mathcal{L}_2\{f*g\} = F_2(s)\ G_2(s)$ where $f*g = \int_{-\infty}^{\infty} f(x)g(t-x)\ dx$

and F_2 and G_2 have regiuns of convergence σ_{f1}, σ_{f2} and σ_{g1}, σ_{g2}.

3. Time multiplication

For $F_2 * G_2 = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F_2(\beta)\ G_2(s-\beta)\ d\beta$

where $\sigma_{f1} < c < \sigma_{f2}$ and $\sigma_{f1} + \sigma_{g1} < \mathrm{Re}\{s\} < \sigma_{f2} + \sigma_{g2}$, ther $\mathcal{L}_2[fg] = F_2 * G_2$

Similar theorems for one sided transform only require appropriate changes in limits.

4. Initial value

$$f(o+) = \lim_{t \to o} f(t) = \lim_{s \to \infty} sF(s)$$

5. Final value

$$f(\infty) = \lim_{t \to \infty} f(t) = \lim_{s \to o} sF(s)$$

where F(s) has poles only in open LHP

E. Table of one-sided Laplace Transforms

Table 2 E.1

f(t)	F(s)		
1. $u_o(t)$	1	,	$\lvert\sigma\rvert < \infty$
2. $u_1(t)$	$1/s$	,	$\sigma > 0$
3. $\epsilon^{-at} u_1(t)$	$\frac{1}{s+a}$	,	$\sigma > -a$
4. $\sin \omega_o t \, u_1(t)$	$\frac{\omega_o}{s^2+\omega_o^2}$	,	$\sigma > 0$
5. $\cos \omega_o t \, u_1(t)$	$\frac{s}{s^2+\omega_0^2}$	,	$\sigma > 0$

3. The Discrete Fourier Transform

The Fourier Transform and the Laplace Transform discussed in Sections I and II, respectively, where defined on continuous time functions. We now want to consider similar transforms defined on discrete time functions, say f(n), which are defined on the integers. In some situations, such functions arise quite naturally, for example in synchronous sequential circuits. A more pertinient example would be a uniformly sampled data system. Here a sampling period T>0 is chosen and the process then selects function values at the instants t_n = nT, where n runs over the set of integers.

A. The Discrete Fourier Transform is defined as

$$\mathcal{F}[f(n)] = F(\gamma) = \sum_{n=-\infty}^{\infty} f(n)\gamma^{-n} \tag{3.1}$$

with the inverse

$$\mathcal{F}^{-1}[F(\gamma)] = \frac{1}{2\pi j} \oint_{\Gamma} F(\gamma)\gamma^{n-1} d\gamma = f(n) \tag{3.2}$$

where $\gamma = e^{j\theta}, -\pi < \theta \leq \pi$, and Γ is the unit circle.

Since we are mostly interested in the Z Transform, the complete analogy between the Discrete Fourier Transform and the Fourier Transform of Section 1 will not be given here. It sufficies to note that (3.1) and (3.2) form a unique pair, and a short table of Discrete Fourier Transforms is given in Table 3.A.1.

As in the continuous signal case, it is necessary to define the discrete singularity functions. For our purposes, consider the "backward difference operator", i.e. the first backward difference operator Δb is defined as $\Delta_b[f(n)] = f(n) - f(n-1)$. Hence starting with the definition

$$\nu_1(n) = \begin{cases} 1, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

i.e. $\nu_1(n)$ is the discrete unit step, then

$$\nu_o(n) = \Delta_b[\nu_1(n)] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

i.e. $\nu_o(n)$ is the discrete unit sample. For the inverse, consider $\Delta_b^{-1}[\nu_o(n)] = \sum_{k=-\infty}^{n} \nu_o(k) = \nu_1(n)$ for all $n \geq 0$. Hence, we may define the discrete singularity functions as follows

$$\vdots$$

$\nu_{-1}(t) = \Delta_b\{\nu o(t)\}$, unit doublet

$\nu o\ (t) =$ unit sample

$\nu_1\ (t) = \Delta_b^{-1}\{\nu o(t)\}$, unit sample

$$\vdots$$

where $\Delta_b \cdot \Delta_b^{-1}$ is the invariant operation.

A short table of transforms is given in Table 3.A.1

Table 3.A.1

	$f(n)$	$F(\gamma) = F(\epsilon^{j\theta})$
1.	$v_o(n)$	1
2.	$v_1(n)$	$\pi u_o(\gamma-1) + \frac{\gamma}{\gamma-1}$
3.	K	$2\pi K u_o(\gamma-1)$
4.	$\epsilon^{j\lambda_o n}$	$2\pi u_o(\gamma-\gamma_o)$, $\gamma_o = \epsilon^{j\lambda_o}$
5.	$\cos \lambda_o n$	$\pi[u_o(\gamma-\gamma_o) + u_o(\gamma-\gamma_o{*})]$

6. $a^n v_1(n)$, $0 < \|a\| < 1$	$\frac{\gamma}{\gamma - a}$
7. $b^{\|n\|}$, $0 < \|b\| < 1$	$\frac{(b^2-1)}{b} \quad \frac{\gamma}{(\gamma-b)(\gamma-b^{-1})}$

A. The Two-Sided (Bilateral) Ƶ-Transform

For the sequence $\{f(n)\}_{n=-\infty}^{\infty}$, this transform is defined as the Laurent series

$$\sum_{n=-\infty}^{\infty} f(n)z^{-n} = \mathcal{Z}_2\{f(n)\} = F_2(z) \qquad (4.1)$$

where $z = |z|\ e^{j\theta}$, and the sum converges uniformly for $0 \leq r_1 < |z| < r_2$, an annulus about $z=0$.

Clearly this is an extension of the Discrete Fourier Sum (3.1) by inclusion in the summand the "damping factor" $/z/$ so that the $\gamma = e^{j\theta}$ in the summand of (3.1) is replaced by $/z/\gamma = /z/e^{j\theta} = z$ in the summand of (4.1)

The Inverse Ƶ transform is

$$f(n) = \mathcal{Z}_2^{-1}\{F_2(z)\} = \frac{1}{2\pi j} \oint_C F_2(z)\, z^{n-1}\, dz \qquad (4.2)$$

here C is a closed contour enclosing $z=0$ in the annulus of convergence of (4.1). Just as in the case of the Inverse Laplace Transform (2.2), the region of convergence of $F_2(z)$ must be considered to determine f(n). For the usual cases of signals that are bounded for large $/n/$, poles of $F_2(z)$ inside the unit circle Γ are associated with $n>0$) and poles of $F_2(z)$ outside Γ are associated with $n<0$.

Theorem 4-1. The conditions

1. f(n) real, n integer .

2 $\sum_{n=-\infty}^{\infty} |f(n)|z|^{-n}| < \infty$, for some $|z|>0$

imply

a. $F_2(z)$ exists

b. $f(n) = \mathcal{Z}_2^{-1}\{F_2(z)\}$.

A short table of transforms is given in Table 4.A.1.

Table 4.A.1

f(n)	$F_2(z)$	
1. $b^{\|n\|}, 0<\|b\|<1$	$\frac{b^2-1}{b} \frac{z}{(z-b)(z-b^{-1})}$,	$\|b\|<\|z\|<\|b^{-1}\|$
2. $v_1(n+k) - v_1(n-k)$	$\frac{z(z^k - z^{-k})}{(z-1)}$,	$\|z\|<\infty$
3. $v_1(n+k) - v_1(n-k-1)$ (symmetric)	$\frac{z(z^k - z^{-(k+1)})}{z-1}$,	$\|z\|<\infty$

b. The One-Sided (Unilateral) $\mathcal{Z}$ Transform

This transform is used for functions $f(n)$ where $f(n) = 0$ for $n<0$. This is emphasized by using the notation $f(n)\nu_1(n)$. The transform is defined as

$$F(z) = \mathcal{Z}\{f(n)\nu_1(n)\} = \sum_{n=0}^{\infty} f(n)\, z^{-n} \tag{4.3}$$

where convergence is for $|z|>r>0$, i.e. a Taylor series about $z^{-1} = 0$. The Inverse Transform is $f(n)\nu_1(n) = \mathcal{Z}^{-1}\{F(z)\} = \frac{1}{2\pi j} \oint_C F(z)\, z^{n-1} dz$ (4.4)

where C is a closed contour enclosing $z=0$ in the region of convergence. A unique transform pair is formed by (4.3) and (4.4).

Theorem 4-2. The conditions

1. $\sum_{n=0}^{N} |f(n)| \le K < \infty$, for some $N>0$

2. $|f(n)| \le Mr^{n}$, for some $M>0$ and $n>N$ implies

$f(z) = \sum_{n=0}^{\infty} f(n)\, z^{-n}$ converges absolutely (and uniformly for $|z|>r$.

Theorem 4-3. The conditions F(s) converges uniformly $|z| > r$

implies

$$f(n) = \frac{1}{2\pi j} \oint_c F(z)\, z^{n-1}\, dz; \; n \geq 0, \text{ where c is as in } (4.4)$$

The Inversion Theorem 4-3 is the general method of finding inverse Ƶ Transforms, using the calculus of residues. Tables may also be used, because of the imigreness of the transform pair. Of special usefulness in this regard is the partial fraction expansion of F(z)/z, to be explained. Also, for F(z) rational, a series expansion of F(z) about $z = \infty$ in terms of z^{-1} (convergent for $/z/ > r$) can be used. The coefficient of z^{-n} in the expansion is f(n). This expansion is easily obtained by the use of long division.

Partial fraction method

Let F(z) = z N(z)/D(z), where N (z) and D(z) are polynomials such that r = degree of D > degree of N, then

1. Simple pole case

$$F(z)/z = \sum_{j=1}^{r} \frac{Aj}{z - \lambda j} \quad , \text{ where the } \lambda j \text{ are}$$

simple zeros of D(z) and

$$Aj = \frac{N(z)}{D(z)} (z - \lambda j) \Big|_{z = \lambda j}$$

2. Multiple Pole Case

$$F(z)/z = \sum_{j=1}^{k} \sum_{i=1}^{mj} \frac{A_{ij}}{(z - \lambda j)} \quad , \text{ where } mj \text{ is the}$$

order of the zero of D(z) at z = λj, and

$$r = \sum_{j=1}^{k} mj \quad , \text{ and}$$

$$A_{ij} = \frac{1}{(m_j - i)!} \frac{d^{m_j - i}}{dz^{m_j - i}} \left[\frac{N(z)(z - \lambda j)^{mj}}{D(z)} \right] \Bigg|_{z = \lambda_j}$$

Then the inverse transforms are, respectively,

1. $f(n) = \sum_{j=1}^{r} A_j (\lambda_j)^n \quad , n \geq 0$

2. $$f(n) = \sum_{j=1}^{k} \sum_{i=1}^{m_j} A_{ij} (\lambda_j)^{n-i+1} \frac{n!}{(i-1)!\,(n-i+1)!} \quad , \text{for } n \geq 0$$

Use of Calculus of Residues

For F(z) meromorphic, thenby use of Cauchy Residue Theorem

$$\sum \text{residues of } F(z) z^{n-1} \Big|_{\text{poles of } F(z) z^{n-1}}$$

$$= \frac{1}{2\pi j} \oint_c F(z) z^{n-1} dz = f(n) \qquad , n \geq 0$$

where c is as in Theorem 4-3.

The inverse Two-Sided Transform can be found using arguments similar to that as in Section 2B, but special care must be taken in defining the contour C since, for $|n| \to \infty$ an essential singularity appears in the integrand of (4.2) due to the factor z^{n-1}.

c. Properties of One sided Transform

1. Linearity

$$\mathcal{Z}[a f(n) + b g(n)] = a F(z) + b G(z)$$

2. Backward difference of f(n)

$$\mathcal{Z}[\Delta_b [f(n)]] = (1 - z^{-1}) F(z) - f(-1)$$

3. "Inverse"of backward difference of f(n)

$$\mathcal{Z}[\Delta_b^{-1} [f(n)]] = \mathcal{Z}\Big[\sum_{k=-\infty}^{n} f(k)\Big] = \frac{1}{1-z^{-1}} F(z) + \frac{1}{1-z^{-1}} \sum_{k=-\infty}^{-1} f(k)$$

4. Integer shift

$$\mathcal{Z}[f(n-k)] = \begin{cases} z^{-k} F(z) & , k > 0 \\ z^{k} F(z) - z^{k} \sum_{j=0}^{k-1} f(j) z^{-j} , & k < 0 \end{cases}$$

5. Frequency shift

$$\mathcal{Z}[f(n) a^n] = F(a^{-1} z)$$

6. Integer multiplication

$$\mathcal{Z}[n f(n)] = z^{-1} \frac{d}{dz^{-1}} F(z)$$

D. Theorems

1. Sum Square

If f(n) satisfies Theorem 4-1 for /z/= 1, then

$$\sum_{n=-\infty}^{\infty}[f(n)]^2 = \frac{1}{2\pi j}\oint_{\Gamma} F_2(z)F_2(z^{-1})z^{-1}dz$$

where Γ is the unit circle.

2. Real convolution

$$\mathcal{Z}[f*g] = F_2(z)G_2(z) \quad , \text{where} \quad f*g = \sum_{k=-\infty}^{\infty} f(k)g(n-k)$$

3. Real multiplication

For $F_2 * G_2 = \frac{1}{2\pi j}\oint_C F_2(\beta)G_2(z-\beta)\beta^{-1}d\beta$

where C is a closed contour enclosing z = 0 in the region of convergence, then

$\mathcal{Z}_2\,[fg] = F_2 * G_2$.

4. Initial value

$$f(o) = \lim_{z\to\infty}(1-z^{-1})\,F(z)$$

5. Final Value

$$f(\infty) = \lim_{z\to 1}(1-z^{-1})\,F(z)$$

where $(1-z^{-1})\,F(z)$ has poles only inside the unit circle.

E. Table of One Sided Z Transforms

Table 4.E.1

	f(n)	F(z)		
1.	$v_0(n)$	1	,	$\lvert z\rvert < \infty$
2.	$v_1(n)$	$\frac{z}{z-1}$	,	$\lvert z\rvert > 1$
3.	$a^n v_1(n)$	$\frac{z}{z-a}$	,	$\lvert z\rvert > \lvert a\rvert$

4. $\sin \lambda_0 n \; v_1(n)$	$\frac{z \sin \lambda_0}{z^2-2z \cos \lambda_0+1}$, $\|z\| > 1$
5. $\cos \lambda_0 n \; v_1(n)$	$\frac{z(z-\cos \lambda_0)}{z^2-2z \cos \lambda_0+1}$, $\|z\| > 1$

F. Correspondence of Theorems

Note that the $\mathcal{Z}$ Transform Theorems 4-1, 4-2, and 4-3 correspond to the Laplace Transform Theorems 2-1, 2-2, and 2-3 respectively. The Fourier Transform Theorem 1-1 has a corresponding Discrete Fourier Transform Theorem (Section 3), but this was not pursued for reasons already mentioned in Section 3.

5. Relationships Among Transforms

There are situations where it is convenient to change from one transform of a function $f(t)$ or $f(n)$ to another transform. One is in finding amplitude spectra from Laplace or $\mathcal{Z}$ Transforms by converting to the appropriate Fourier Transform. A second is that the computation of two-sided transforms from known one-sided transforms is sometimes easy. A third is that certain operations on signals in the time domain, e.g. sampling and truncating, may be done equivalently by operations on the signal transforms.

Finally, looking at the various transforms as one ball of wax should aid in the understanding of continuous time and discrete time systems.

A. General Relationships

1. Two common relationships are between $F(s) = \mathcal{L}[f(t)]$ and $F(j\omega) = \mathcal{F}[f(t)]$ and between $F(z) = \mathcal{Z}[f(n)]$ and $F(\gamma) = \mathcal{F}_d[f(n)]$.

   a. For the transforms of $f(t)$, if the poles of $F(s)$ lie in the LHP, the substitution $s = j\omega$ or vice versa relates the Fourier and Laplace Transforms.

   b. For the transforms of $f(n)$, if the poles of $F(z)$ lie in Γ , the substitution $z = e^{j\theta}$ or vice versa relates the Discrete Fourier and $\mathcal{Z}$ Transforms.

As examples, compare Table 1.E.1 (entry 2) with Table 2 E.1. (entry 3), and compare Table 4.E.1 (entry 3) with Table 3.A.1(entry6)

2. Similarly, there are relationships between

$F_2(s) = \mathcal{L}_2[f(t)]$ and $F(j\omega) = \mathcal{F}[f(t)]$ and between $F_2\ (z) = \mathcal{Z}_2[f(n)]$ and $F(\gamma) = \mathcal{F}_d[f(n)]$.

a. For the transforms of f(t), if the poles of $F_2(s)$ lie in LHP or RHP, the substitution $s= j\omega$ or vice versa relates the Fourier and the Bilaterial Laplace Transforms.

b. For the Transforms of f(n), if the poles of $F_2(z)$ lie inside Γ or outside Γ, the substitution $z = e^{j\theta}$ or vice versa relates the Discrete Fourrier and the Bilaterial Z - Transforms.

As examples, compare Table 1.E.1(3) with Table 2.A.1(1), and compare Table 4.A.1 (1) with Table 3.A.1 (7). The above substitutions also occur as trivial cases when $F_2\ (s)$ converges only for $\sigma=0$ or $F_2(z)$ converges only for $/z/ = 1$. For example, look at Table 1.E.1(5) and Table 2.A.1(4).

3. Special cases are simple poles of F(s) on Δ(i.e.$s=j\omega$) or simple poles of F(z) on Γ.

a. For transforms of f(t), if the poles of F(s) are simple poles at $s=s_m, S_m \in \Delta$, then

$$\mathcal{F}[f(t)] = F(s)\Big|_{s=j\omega} + \pi\sum_m c_m u_o(\omega-\omega_m)$$

where cm is the residue of F(s) in the pole $s = s_m$

b. For the transforms of f(n), if the poles of F(z) are simple poles at $z = z_m, z_m \in \Gamma$, then

$$\mathcal{F}_d[f(n)] = F(z)\Big|_{z=\gamma} + \pi\sum_m c_m u_o(\gamma - e^{\pm j\theta_m})$$

where cm is the residue of F(z) in the pole $z= z_m= e^{\pm j\theta_m}$.

As examples, compare Table 2.E.1 (5) with Table 1.E.2(6), and compare Table 4.E.1(2) with Table 3.A.1(2).

4. When f(t) or f(n) is periodic the appropriate Fourier Transform has a simple relationship to the appropriate Fourier Series (the Fourier Series of continous time or discrete time periodic signals was not covered in the preceeding material).

a. For f(t) periodic with period To,

$$\mathcal{F}[f(t)] = 2\pi\sum_{k=-\infty}^{\infty} c_k u_o(\omega - k\omega_o)$$

where $\omega o = {}^{2\pi}/_{T_o}$ and the c_k are the coefficients of the Fourier Expotential Series of f(t).

b. For f(n) periodic with period N_o (N_o must be a positive integer),

$$\mathcal{F}_d[f(n)] = 2\pi\sum_{k=-\infty}^{\infty} c_k u_o(\gamma - \gamma_o{}^k)$$

where $\lambda_o = \frac{2\pi}{N_o}$ and $\gamma_o = e^{j\lambda o}$ and the c_k are the coefficients of the Discrete Fourier Exponectial Series of f(n).

As examples see Table 1.E.2(4) for the continuous periodic signal f(t) and see Table 3.A.1(5) for the discrete periodic signal f(n).

B. Unilaterial and Bilaterial Laplace and Z Transforms.

The two-sided Laplace Transform is

$$\mathcal{L}_2[f(t)] = F_2(s) = \int_{-\infty}^{\infty} f(t) e^{-st} dt$$

$$= \int_{-\infty}^{0} f(t) e^{-st} dt + \int_{0}^{\infty} f(t) e^{-st} dt$$

$$= \int_{0}^{\infty} f(-\tau) e^{s\tau} d\tau + \mathcal{L}[f(t)]$$

$$= \mathcal{L}[f(-t)]\Big|_{s \to -s} + F(s)$$

where we have assumed f(t) satisfies Theorem 2-1. Note the transformations on the right hand side can be done using a table of one-sided transfroms, and also that special care must be taken when singularity functions $u_j(t)$, $j \leq 0$, occur. For the special case of f(t) an even function it follows that

$F_2(s) = F(-s) + F(s)$
where $F(s) = \mathcal{L}[f(t)]$.

Similarity the two sided Z-Transform gives

$$\mathcal{Z}_2[f(n)] = F_2(z) = \mathcal{Z}[f(-n)]\Big|_{z \to z^{-1}} + \mathcal{Z}[f(n)] - f(0)$$

where F(z) = [f(n)]. For the special case of f(n) an even function, then

$F_2(z) = F(z^{-1}) + F(z) - f(0)$.

C. Truncation of f(t) or f(n).

In order to assure the physical realizability of a given unit impulse response h(t) or a given unit pulse response k(n), it may be necessary (from causality) to make the value of the function identically zero for negative arguements. That is, replace h(t) by h(t) $u_1(t)$ or k(n) by k(n) $\nu_1(n)$. Hence if the transforms $H_2(s)$ or $K_2(z)$ are known, the transforms H(s) or $K_2(z)$ include the imaginary axis of the s-plane and the boundary of the unit circle in the z-plane, respectively, then:

$H(s) = \left[H_2(s)\right]_+$

where the right hand side means expand $H_2(s)$ in partial fractions and retain only those terms with poles in the LHP, i.e. with a time domain response nonzero only for o<t,

$$K(z) = \left[K_2(z)\right]_+$$

where the right hand side means expand $K_2(z)/z$ in partial fractions, retain only those terms with poles inside the unit circle, and multiply those terms by z. These final terms have a time domain response non zero only for 0 n.

As examples, from Table 2.A. (1),

$$F_2(s) = \frac{2a}{a^2 - s^2} = \frac{-2a}{(s+a)(s-a)} \quad , 0<a$$

so in partial fractions

$$F_2(s) = \frac{1}{s+a} - \frac{1}{s-a} \quad \text{and,}$$

$$F(s) = \left[F_2(s)\right]_+ = \frac{1}{s+a}$$

and note this cehcks Table 2.E.1 (3). Similarly, from Table 4.A.1(1),

$$F_2(z) = \frac{b^2-1}{b}\,\frac{z}{(z-b)(z-b^{-1})} \quad , 0<|b|<1 \ , \text{so}$$

$$\frac{F_2(z)}{z} = \frac{1}{z-b} - \frac{1}{z-b^{-1}} \quad \text{and}$$

$$F(z) = \left[F_2(z)\right]_+ = z/z-b$$

and note this checks Table 4.E.1(3).

D. Relationship between continous time and discrete time transforms; uniform impulse sampling.

Let T>0 and define

$$u_T(t) = \sum_{n=0}^{\infty} u_0(t-nT) \quad , t \geq 0^-$$

a sequence of unit impulses spaced T seconds apart starting at the origin and continuing to the right. Let $g(t) = 0$, $t<0$, and let

$$g^*(t) = g(t)\,u_T(t) = \sum_{n=0}^{\infty} g(nT)\,u_0(t-nT)$$

denote the uniformly impuse sampled g(t). Then

$$\mathcal{L}[g^*(t)] = G^*(s) = \sum_{n=0}^{\infty} g(nT)\,e^{-nTs} \quad \text{Now let } z = e^{Ts}, \text{ then}$$

$$G^*(s)\Big|_{z=e^{Ts}} = \sum_{n=0}^{\infty} g(nT)\,z^{-n} = \mathcal{Z}[g(nT)]$$

Clearly, if T=1, this is $G(z)=\mathcal{Z}[g(n)]$, but the Z Transforms already discussed may be used provided T is included appropriately.

For example, consider

$$g(t) = e^{-bt}u_1(t) \quad , 0<b \quad \text{Then,}$$

$$g(nT) = e^{-bnT}u_1(nT) \quad , \text{and}$$

$$\mathcal{Z}[g(nT)] = \sum_{n=0}^{\infty} e^{-bnT} z^{-n} = \frac{z}{z - e^{-bT}}$$

Now consider the discrete time function of Table 4.E.1(3),

$$f(n) = a^n v_1(n) \quad , \quad 0 < a < 1 \quad , \text{and}$$

$$F(z) = \mathcal{Z}[f(n)] = \frac{z}{z - a} \quad , \quad |z| > |a|$$

Thus $g(nT) = f(n)$ and $\mathcal{Z}[g(nT)] = F(z)$ for $e^{-bt} = a$, so the notation $G(z) =$. $\mathcal{Z}[g(nt)] = G^*(s)\big|_{z = e^{sT}}$ should cause no difficulty.

REFERENCES

1. Lewis, J. B., "Notes on System Theory", The Pennsylvania State University, 1972.

2. Papoulis, A., "The Fourier Integral and Its Applications", McGraw-Hill Book Co., New York, 1962.

3. Bracewell, R., "The Fourier Transform and It's Applications", McGraw-Hill Book Co., New York, 1965.

4. Le Page, W. "Complex Varrables and the Laplace Transform for Engineers", McGraw-Hill Book Co,, New York, 1961.

5. Schuarz, R. and Friedland, B., "Linear Systems", McGraw Hill Book Co., New York 1965.

6. DeRusso, P., et.al., "State Varrables for Engineers", J. Wiley and Sons, Inc., New York 1965.

7. Freeman, H., "Discrete-Time System", J. Wiley and Sons, Inc., New York, 1965.

Part II. A Review of Methods of Generating Discrete Systems from Prototype Continuous Systems

Introduction

There are two basic approaches to modelling a continuous time process by a discrete time process. One approach is to start with the continuous time input-output mathematial relationships and to approximate these with descrete time input-output mathematial relationships. For example, an ordinary, linear, constant coefficient differential equation might be approximated by an ordinary, linear, constant cofficient difference equation. A second approach is to design a continuous time system (thus fixing the topology of the system) and to approximate this existing continuous time system (prototype) with a discrete time system.

This second approach is often taken because of the wealth of design concepts that exist for continuous time systems, i.e. many people feel more at ease using continuous time system design ideas. In general however, this approach has two approximations inherent in it. One is in the process of obtaining realizable transfer functions from which to proceed and the other is in approximating the realized continuous system by a discrete system. Hence, the first approach is the more fundamental of the two.

The frequency domain approach for CT, L, TI systems and DT, L, TI systems was discussed previously in [14], i.e., the Fourier Transform and the Laplace Transform for the former, and the Discrete Fourier Transform and the Z-Transform for the latter. Relationships among these transforms were also obtained.

The time domain approach for these systems should be briefly described. In the CT, L, TI case, the input-output relationship for $t_0 < t$ is the differential equation

$$\sum_{i=0}^{N} d_i D^i y(t) = \sum_{i=0}^{M} c_i D^i u(t) \qquad (1)$$

where $D^i = \frac{d^i}{dt^i}$, along with appropriate boundary conditions. The general solution is the sum of a homogeneous solution and a particular solution. In the DT, L, TI case, the input-output relationship for $n_0 < n$ is the difference equation

$$\sum_{i=0}^{N} b_i E^{-i} z(n) = \sum_{i=0}^{M} a_i E^{-i} v(n) \qquad (2)$$

where $E^{-i} f(n) = f(n-i)$, along with appropriate boundary conditions. The general solution is the sum of a homogeneous solution and a particular solution.

More directly however, the difference equation provides an explicit relationship between the input and output which can be seen by rewriting (2) as

$$z(n) = -\sum_{i=1}^{N} \frac{b_i}{b_o} z(n-i) + \sum_{i=o}^{M} \frac{a_i}{b_0} v(n-i) \qquad (3)$$

Thus the nth value of the output can be computed from the nth input value and the N and M past values of the output and input, respectively. Hence the difference equation not only represents the system for theoretical purposes, but it may also serve as a computational realization of the system.

For a unit sample (pulse) input, $v(n) = v_0(n)$, let the response be $z(n) = k(n)$, i.e., $k(n)$ is the unit sample response. Because of the properties of some digital processing techniques, it is useful to distinguish between two classes of DT systems v_{iz} those for which $k(n) \neq o$ only for a finite duration and those for which $k(n) \neq o$ for an infinite duration. The former are referred to as finite impulse response (or recursive systems and the latter are referred to as infinite impulse response (or recursive) systems. Since the various systems we are interested in will be used for their signal processing properties, we will refer to them as filters. From the previous paragraph, the recursive filter can be thought of as having infinite memory. On the other hand, the nonrecursive filter has finite memory.

For CT systems and for a unit impulse input, $u(t) = u_0(t)$, i.e.,$h(t)$ is the unit impulse response. (So $k(n)$ for DT systems corresponds to $h(t)$for CT systems.) Then the forced response of a CT system to $u(t)$ is given by the convolution integral

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau)u(t-\tau)d\tau \\ &= \int_{-\infty}^{\infty} h(t-\tau)\,u(\tau)\,d\tau \end{aligned} \qquad (4)$$

Similarly the forced response of a DT system to $v(n)$ is given by the convolution sum

$$\begin{aligned} z(n) &= \sum_{m=-\infty}^{\infty} k(m)\, v(n-m) \\ &= \sum_{m=-\infty}^{\infty} k(n-m)\, v(n) \end{aligned} \qquad (5)$$

Now if N=o in (2) so that (3) becomes

$$z(n) = \sum_{i=o}^{M} \frac{a_i}{b_o} v(n-i) \qquad (6)$$

then it corresponds to a nonrecursive system, and by comparing (6) with (5) we see that for such a system the unit sample response is

$$k(n) = \begin{cases} \frac{a_n}{b_0} & , n = 0,1,2,3,\cdots,M \\ 0 & , \quad \text{else} \end{cases}$$

and the finite duration is evident. In contrast, for a recursive system, N in (2) must be greater than zero (and usually M $\leq$ N)

As is well know, the transfer function of the continuous filter can be obtained from the differential equation (1) by making a correspondence between the continuous differential operator D and the Laplace Transform variable s, thus

$$H(s) = \frac{Y(s)}{U(s)} = \frac{\sum_{i=0}^{M} c_i s^i}{\sum_{i=0}^{N} d_i s^i} \qquad (7)$$

Similarly, the transfer function of the discrete filter can be obtained from the difference equation (2) by making a correspondence between the discrete unit advance operator E and the Ƶ - Transform variable z, thus

$$K(z) = \frac{Z(z)}{V(z)} = \frac{\sum_{i=0}^{M} a_i z^{-i}}{\sum_{i=0}^{N} b_i z^{-i}}$$

$$= \frac{\sum_{i=0}^{M} a_i z^{-i}}{1 + \sum_{i=1}^{N} b_i z^{-i}} \qquad (8)$$

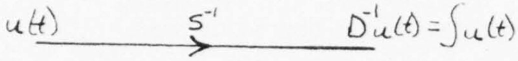
where $b_0 = 1$ can be assumed without loss of generality.

1. Discrete Filter Structures

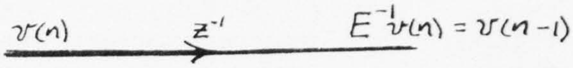
There are well known canonic form networks for continuous filters employing integrators, multipliers, and summers. It is useful to have similar structures for discrete filters. In order to employ a saving of space, these are best given in flow graph form. The correspondance between the continuous filter elements and discrete filter elements are shown in the following two columns.

Continuous Elements

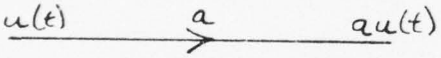
Discrete Elements



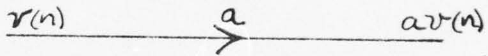
Integrator



Unit Delay

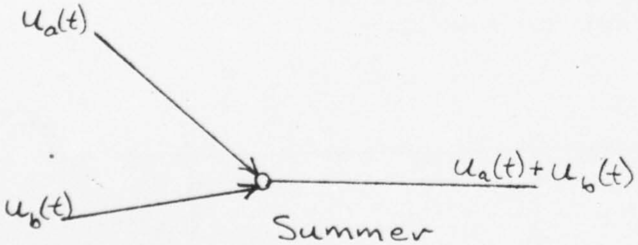


Multiplier

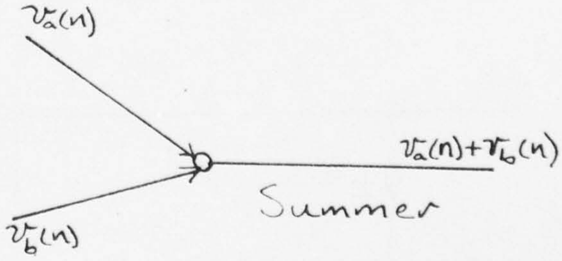


Multiplier

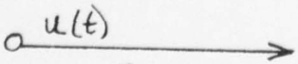
(when a=1, we will simply use $\longrightarrow$)



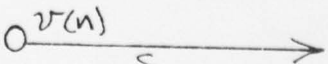
Summer



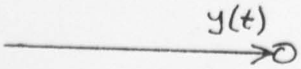
Summer



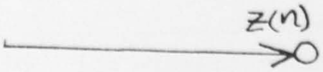
Source



Source



Sink



Sink

It should be clear that for each rational transfer function, there corresponds a variety of different network configurations. For the digital filter case, the choice between these different realizations is computational complexity. This is from the fact that storage for the variables and constants, as well as means for delaying, multiplication, and addition must be provided. In general, multiplication is a time-consuming operation and each delay element requires the use of a memory register, thus structures with the minimum number of these elements are often desirable. Yet the effects of finite register length in actual hardware realizations depend on the structure chosen.

2. Recursive Filter Structures

Recalling the material discussed in the Introduction, we see that a rational transfer function of the form

$$K(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 - \sum_{k=1}^{N} a_k z^{-k}} \tag{2.1}$$

corresponds to the difference equation

$$z(n) = \sum_{k=1}^{N} a_k z(n-k) + \sum_{k=0}^{M} b_k v(n-k) \tag{2.2}$$

A direct flow graph realization of this filter can be written from (2.2). Assuming M=N in (2.2), the network of Fig. 2.1 results.

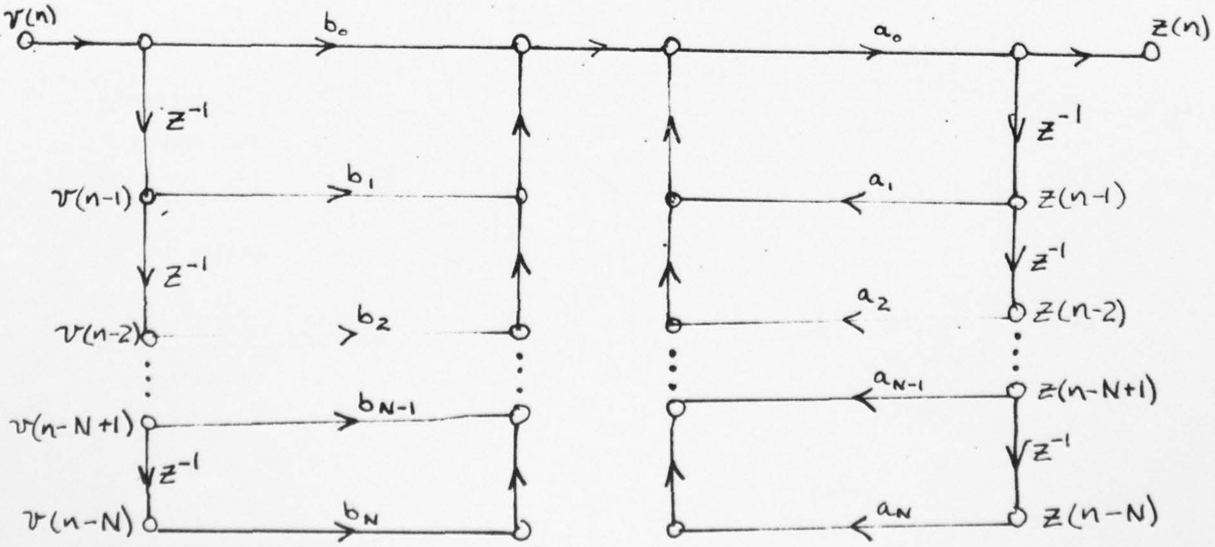


Figure 2.1 Direct form I.

Note that the direct form I realization of Fig. 2.1 has more nodes than are necessary, but that each node has at most two inputs and hence is consistent with the fact that addition in digital hardware typically involves two numbers at a time. Also, in terms of signal flow, we see that Fig. 2.1 first realized the zeros and then the poles of (2.1). Clearly we could reverse this order, i.e., first realize the poles and then the zeros of (2.1). Doing this, then the unit delays can be combined since they have the same input, thus obtaining the network with the minimum number of delays shown in Fig 2.2.

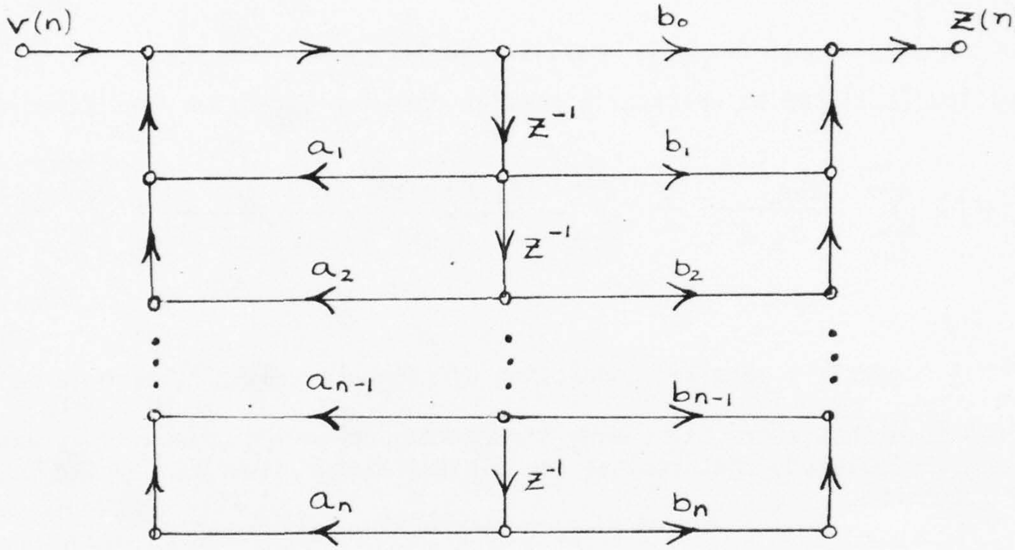


Figure 2.2 Direct form II.

Many other networks with the minimum number of delays can be obtained.

Equation (2.1) can be written in the factored form

$$K(z) = A \frac{\prod_{k=1}^{M_1} (1 - g_k z^{-1}) \prod_{k=1}^{M_2} (1 - h_k z^{-1})(1 - h_z^* z^{-1})}{\prod_{k=1}^{N_1} (1 - e_k z^{-1}) \prod_{k=1}^{N_2} (1 - d_k z^{-1})(1 - d_k^* z^{-1})} \tag{2.3}$$

Thinking of (2.3) composed of factors of the general form

$$K_k(z) = \frac{1 + \beta_{1k} z^{-1} + \beta_{2k} z^{-2}}{1 - \alpha_{1k} z^{-1} - \alpha_{2k} z^{-2}} . \tag{2.4}$$

a direct form II realization of (2.4) is shown in Figure 2.3,

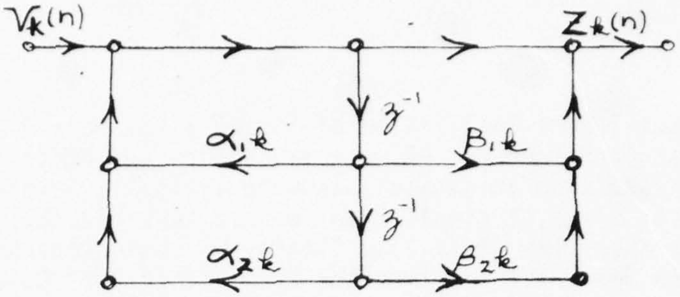


Figure 2.3 A second order realization. nd a cascade of such networks realizes (2.3).

and a cascade of such networks realizes (2.3).

Equation (2.1) can be written in partial fraction expansion form from (2.3) as

$$K(z)=\sum_{k=1}^{N_1}\frac{A_k}{1-c_k z^{-1}}+\sum_{k=1}^{N_2}\frac{B_k(1-e_k z^{-1})}{(1-d_k z^{-1})(1-d_k^* z^{-1})}+\sum_{k=0}^{M-N}C_k z^{-k} \qquad (2.5)$$

Clearly this suggests a parallel connection of networks realizing each term.

We see then that there exist many structures, depending on how the transfer function is manipulated, that realize the digital filter discribed by the transfer function.

3. Nonrecursive Filter Structures

From the Introduction, the nonrecursive filter has the transfer function

$$K(z)=\sum_{n=0}^{M}k(n)z^{-n} \qquad (3.1)$$

and from this, just as in Section 2, many network realizations can be obtained. We will not bother to repeat the work here.

4. Discrete Filters from Continuous Filters.

From the Introduction, we recall that a continuous filter can be described by the differential eauation

$$\sum_{k=0}^{N}c_k D^k y(t) = \sum_{k=0}^{M}d_k D^k u(t) \qquad (4.1)$$

where $D^k f(t) = \frac{d^k f(t)}{dt^k}$, or by the transfer function

$$H(s) = \frac{Y(s)}{U(s)} = \frac{\sum_{k=0}^{M} d_k s^k}{\sum_{k=0}^{N} c_k s^k} \tag{4.2}$$

and the forced response can be found by convolution as

$$y(t) = \int_{-\infty}^{\infty} u(\tau)\, h(t-\tau)\, d\tau \tag{4.3}$$

The continuous filter is assumed to be causal (no output before the application of an input), stable (poles in LHP with at most simple poles on $j\omega$), and satisfies some magnitude and/or phase requirements as a function of real frequency ω.

Similarly, a discrete filter can be described by the difference equation

$$\sum_{k=0}^{N} a_k E^{-k} z(n) = \sum_{k=0}^{M} b_k E^{-k} v(n) \tag{4.4}$$

where $E^{-k}f(n) = f(n-k)$, or by the transfer function

$$K(z) = \frac{Z(z)}{V(z)} = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}} \tag{4.5}$$

and the forced response can be found by convolution as

$$z(n) = \sum_{i=-\infty}^{\infty} v(i)\, k(n-i) \qquad (4.6)$$

Thus, in generating a discrete filter from a continuous prototype filter, either K(z) or k(n) must be obtained while preserving the properties of causality, stability, and desirable magnitude (or phase) characteristics.

TIME-DOMAIN TECHNIQUES

A. Design using convolution

Only nonrecursive filters can be realized exactly with this technique (since from (6) the unit sample response is of finite duration), but recursive filters could be approximated by truncating the infinite duration unit sample response.

Thus assume we have a unit sample response given by the set of M + 1 sample response given by the set of m + 1 samples k(o), k(T), k(2T),....k(M-1)T), k(MT). Then from (6), we have

$$z(nT) = \sum_{m=0}^{M} k(mT)\, v(n-m)T) \qquad (4.A.1)$$

and

$$\begin{aligned} K(z) &= \frac{Z(z)}{V(z)} = \sum_{m=0}^{M} k(mT)\, z^{-m} \\ &= k(o) + k(T)z^{-1} + k(2T)z^{-2} + \cdots + k(MT)z^{-M} \\ &= \frac{k(o)z^{M} + k(T)z^{M-1} + k(2T)z^{M-2} + \cdots + k(MT)}{z^{M}} \end{aligned}$$

(4.A.2)

from (4.A.2) we may factor K(z) as

$$K(z) = \begin{cases} \prod_{i=0}^{M/2} \left(\frac{A_{2i} z^2 + A_{1i} z + A_{0i}}{z^2} \right) & , M \text{ even} \\ \frac{A_1 z + A_0}{z} \prod_{i=0}^{\frac{M-1}{2}} \left(\frac{A_{2i} z^2 + A_{1i} z + A_{0i}}{z^2} \right) & , M \text{ odd} \end{cases} \tag{4.A.3}$$

Now consider the general 2nd order filter described by

$$K_i(z) = \frac{A_{2i} z^2 + A_{1i} z + A_{0i}}{z^2 - B_{1i} z - B_{0i}} \tag{4.A.4}$$

This corresponds to the difference equation

$$\begin{aligned} z(nT) = {} & A_{2i} v(nT) + A_{1i} v((n-1)T) + A_{0i} v((n-2)T) \\ & + B_{1i} z((n-1)T) + B_{0i} z((n-2)T) \end{aligned} \tag{4.A.5}$$

In implementing this algorithm, each output sample z(nt) must be calculated in the sampling period, T seconds. Thus the sampling frequency $\frac{1}{T}$ is limited by the time required to do the computations required in (4.A.5). Thus hardware may be needed to supplement software. Also important is the length of computer word or register length, since these result in noise (resulting in limit cycles or oscillations). These considerations are discussed elsewhere [13].

Finally, the filter described by (4.A.1) or (4.A.2), is realized by cascading 2nd order sections realized as described above from (4.A.3).

B. Design by impulse response matching

This technique consists of choosing the unit sample response of the discrete filter to be the same as equally spaced samples of the unit impulse response of the continuous filter i.e. for a sampling period T,

$$k(n) = h(nT) \tag{4.B.1}$$

Assume the continuous filter has a rational transfer function as in (7)

$$H(s) = \mathcal{L}[h(t)] = \frac{\sum_{i=0}^{M} c_i s^i}{\sum_{i=0}^{N} d_i s^i} \tag{4.B.2}$$

and that $M<N$. Then (4.B.2) can be expanded in partial fractions as

$$H(s) = \sum_{k=1}^{N} \frac{A_k}{s-s_k} \tag{4.B.3}$$

and

$$h(t) = \mathcal{L}^{-1}\left[H(s)\right] = \sum_{k=1}^{N} A_k e^{skt} u_1(t) \quad . \tag{4.B.4}$$

Then by (4.B.1),

$$k(n) = h(nT) = \sum_{k=1}^{N} A_k (e^{skT})^n \, u_1(nT) \tag{4.B.5}$$

and from [14], $A_k(e^{s_kT})^n u_1(nT) = \mathcal{Z}\left[\frac{A_k}{1-e^{s_kT}z^{-1}}\right]$

so from (4.B.5), $K(z) = \mathcal{Z}\left[k(n)\right] = \sum_{k=1}^{N} \frac{A_k}{1-e^{s_kT}z^{-1}}$

$$= \sum_{k=1}^{N} \frac{A_k}{1-z_k z^{-1}} \tag{4.B.6}$$

Comparing (4.B.3) and (4.B.6) we see that a pole s_k in the s-plane maps to a pole z_k in the z-plane according to $z_k = e^{s_kT}$, so stability is preserved since if $\mathrm{Re}[s_k] \leq 0$ then $|z_k| \leq 1$. However, the impulse matching method <u>does not</u> correspond to a mapping of the s-plane to the z-plane since it is easily seen that the zeros are not mapped in the same way as the poles.

This can be seen in another way also.

$$K(z)\Big|_{z=e^{sT}} = \frac{1}{T} \sum_{k=-\infty}^{\infty} H\left(s + j\frac{2\pi k}{T}\right) \tag{4.B.7}$$

and for $s = j\omega/T = j\Omega$

$$K(e^{j\omega}) = \frac{1}{T} \sum_{k=-\infty}^{\infty} H\left(j\frac{\omega}{T} + j\frac{2\pi k}{T}\right) \tag{4.B.8}$$

so the frequency response of the discrete filter will not be the same as that of the continuous filter because of the interference (aliasing) between successive terms on the right hand side of (4.B.8), unless h(t) is a bandlimited signal such that

$$H(j\Omega) = 0 \quad \text{for } |\Omega| \geq \frac{\pi}{T} \tag{4.B.9}$$

for then

$$K(e^{j\omega}) = \frac{1}{T} H\left(j\frac{\omega}{T}\right), \quad |\omega| \leq \pi \tag{4.B.10}$$

So if the continuous filter is "bandlimited" then the discrete filter frequency response will "closely approximate" the continuous filter frequency response, i.e.

$$K(e^{j\omega}) \simeq \frac{1}{T} H\left(j\frac{\omega}{T}\right) \tag{4.B.11}$$

This will result in high gain for small T, and in such a case,

$$K(z) = \sum_{k=1}^{N} \frac{TA_k}{1-e^{s_kT} z^{-1}} \tag{4.B.12}$$

should be used in place of (4.B.6).

This corresponds to

$k(n) = Th(nT)$
in place of (4.B.1) (4.B.13)

Clearly a realization of the discrete filter from (4.B.6) implies a parallel structure but as discussed in Section 2 many other network structures are possible.

EXAMPLE: If M<N in (4.B.2), then the technique described to arrive at (4.B.6) is unambigious. If M=N, a difficulty arises. Consider a lead-lag network

$$H(s) = \frac{as+1}{bs+1} = \frac{\frac{a}{b}s + \frac{1}{b}}{s+\frac{1}{b}} = \frac{c_1 s + c_0}{s + d_0}$$

Then (4.B.3) is of the form

$$H(s) = C_1 + \frac{C_0 - C_1 d_0}{s + d_0} \qquad (4.B.14)$$

Recall that in a realization of this continuous system the constant term (c_1) on the right hand side corresponds to a direct connection (a multiplier of value c_1) between input and output. Now in the discrete system, this term should have the same meaning. Continuing, from (4.B.14) we get

$$h(t) = \mathcal{L}^{-1}[H(s)] = C_1 u_0(t) + (C_0 - C_1 d_0) e^{-d_0 t} u_1(t) \qquad (4.B.15)$$

Then by (4.B.1)

$k(n) = h\ (nT) = c_1 U_0\ (nT) + (co\text{-}c\ d_0)(e^{-dot})^n\ u_1\ (nT)$ (4.B.16)

Now to keep the multiplier invariant, the first term on the right hand side of this equation, $c_1 u_0(nT)$, must be interpreted as $C_1 V_0(nT)$. Doing this, then we get

$$\begin{aligned} K(z) = \mathcal{Z}[k(n)] &= c_1 + \frac{c_0 - c_1 d_0}{1 - e^{-d_0 T} z^{-1}} \\ &= \frac{c_1(1 - e^{-d_0 T} z^{-1}) + (c_0 - c_1 d_0)}{1 - e^{-d_0 T} z^{-1}} \\ &= \frac{ab(1 - e^{-T/b} z^{-1}) + (b-a)}{b^2(1 - e^{-T/b} z^{-1})} \end{aligned} \qquad (4.B.17)$$

This result is different than that reported in [3]. However, the arguments used here to arrive at (4.B.17) are physical and straightforward.

In the same manner, if M=N+1 in (4.B.2), then (4.B.3) is of the form

$$H(s) = H_1 s + H_0 + \sum_{k=1}^{N} \frac{A_k}{s - s_k} \qquad (4.B.18)$$

Again the constant term H_0 on the right hand side corresponds to a multiplier connected between input and output, and we keep this invariant in the discrete system as discussed above. The term $H_1 s$ on the right hand side corresponds to a differentiator connected between input and output, and in the discrete system this could be interpreted as, say, the first backward difference in the form

$$\begin{aligned} z(n) &= H_1 \frac{v(n) - v(n-1)}{T} \\ &= H_1 \frac{1}{T} \Delta_b [v(n)] \end{aligned} \tag{4.B.19}$$

We have the notion here that as the sampling period T is made smaller, the approximation of a continuous derivative, uniformly sampled at instants nT, by the first backward difference divided by T becomes better.

Similarly, for $M > N+1$ in (4.B.2), terms of the form $H_n s^m$ appear on the right hand side of (4.B.18) and in the discrete system these correspond to

$$H_m \frac{1}{T^m} \Delta_b^m [v(n)] = H_m \frac{1}{T^m} \Delta_b \left[\Delta_b^{m-1} [v(n)] \right] \tag{4.B.20}$$

Also, we have

$$\mathcal{Z}\left[H_m \frac{1}{T^m} \Delta_b^m [v(n)] \right] = H_m \frac{1}{T^m} (1 - z^{-1}) V(z) \tag{4.B.21}$$

Practically, however, the case of $M \geq N$ in (4.B.2) discussed above would be handled in a different way. Let the asymptotic behavior of H(s) for large s be of the form

$$\lim_{s \to j\infty} |H(s)| = \lim_{s \to j\infty} \left| \frac{1}{(s/\omega c)^k} \right| \tag{4.B.22}$$

where k is an integer. From (4.B.8), $k(e^{j\omega})$ is periodic in ω of period $\omega s = \frac{2\pi}{T}$ and, according to the discussion following (4.B.8), we see that in the baseband defined by $|\omega| \leq \frac{\omega s}{2}$ the frequency characteristics of the discrete filter differ from those of the continuous filter by the amount of aliasing. If H(s) is bandlimited to the baseband, there is no aliasing.

But for rational H(s), it is not bandlimited. The magnitude of the error resulting from aliasing can be shown to be related to (4.B.22). If k in (4.B.22) is positive and large and $\omega_c << \omega_s/2$ then aliasing is small and good results are obtained. Conversely, if k is not positive and large (e.g. for an elliptic filter k=1, and for the case M>N then k<o) or if $\omega_c \simeq \omega_s/2$ (e.g. wideband situations) then bad results are obtained.

A way out in these cases might be to add in cascade a wideband low-pass guard filter G(s) having a sufficiently large k in the sense of (4.B.22) and with flat magnitude and linear phase characteristics in the frequency range where we want good results. Thus we form the augmented continuous filter

$$H_a(s) = G(s)\,H(s) \tag{4.B.23}$$

and apply the impulse response matching technique to it.

Of course a G(s) with sufficiently large k, flat magnitude, and linear phase is asking a lot. One possibility is to use an all pole low pass filter (e.g. Butterworth of high order can supply the large k and flat magnitude), but such a filter does not have very linear phase. However, we could follow the filter described by (4.B.23) with an all-pass filter designed for phase equalization.

C. Design by matching other singularity responses.

Clearly the idea contained in Section 4.B. can be extended to responses of other singularity functions. For example, if the prototype continuous filter has good step response characteristics, say small rise time and low peak overshoot, we might want to preserve these in the discrete filter. So the design could be based on a step response matching procedure.

Let the continuous filter step response be

$$a(t) = \int_{-\infty}^{t} h(\lambda)\,d\lambda \tag{4.c.1}$$

so that

$$A(s) = \mathcal{L}[a(t)] = \frac{H(s)}{s} \tag{4.c.2}$$

from (4.B.3), then

$$A(s) = \frac{A_0}{s} + \sum_{k=1}^{N} \frac{B_k}{s - s_k} \tag{4.c.3}$$

and

$$a(t) = \mathcal{L}^{-1}[A(s)] = A_0 u_1(t) + \sum_{k=1}^{N} B_k e^{s_k t} u_1(t) \tag{4.c.4}$$

Let $\alpha(n)$ be the discrete filter response to the discrete unit step v_1 (n). We want

$$\begin{aligned}\alpha(n) &= a(nT)\\ &= A_0 u_1(nT) + \sum_{k=1}^{N} B_k (e^{s_k T})^n u_1(nT)\end{aligned} \tag{4.c.5}$$

so

$$\mathcal{Z}[\alpha(n)] = \frac{A_0}{1-z^{-1}} + \sum_{k=1}^{N} \frac{B_k}{1-e^{s_k T} z^{-1}} \tag{4.c.6}$$

Now a network configuration that realizes (4.c.6) may be chosen, as before depending on how (4.c.6) is manipulated.

Finally, cases where (4.c.2) is an improper rational function can be handled as in Section 4.B.

EXAMPLE: Consider the lead-lag example of the previous section.

$$H(s) = \frac{as+1}{bs+1} = \frac{\frac{a}{b}s + \frac{1}{b}}{s+\frac{1}{b}} = \frac{c_1 s + c_0}{s + d_0} \tag{4.c.7}$$

Then the step response is

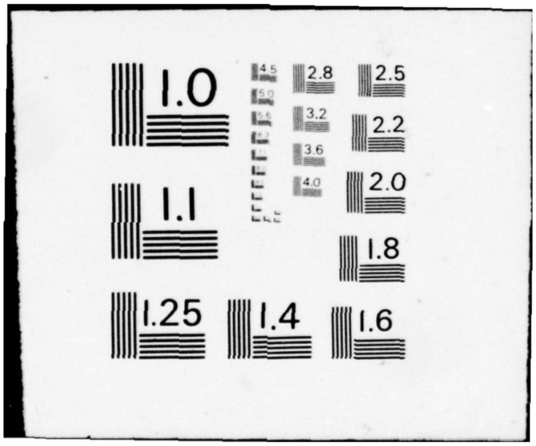
$$A(s) = \frac{H(s)}{s} = \frac{c_1 s + c_0}{s(s+d_0)} \tag{4.c.8}$$

and the partial fraction expansion is

$$\begin{aligned}A(s) &= \frac{c_0/d_0}{s} + \frac{\frac{c_1 d_0 - c_0}{d_0}}{s+d_0}\\ &= \frac{1}{s} + \frac{\frac{a-b}{b}}{s+\frac{1}{b}}\end{aligned} \tag{4.c.9}$$

AD-A031 017 AUBURN UNIV ALA SCHOOL OF ENGINEERING F/G 5/2
1975 USAF/ASEE SUMMER FACULTY RESEARCH PROGRAM.(U)
SEP 75 J F O'BRIEN F44620-75-C-0031
UNCLASSIFIED AFOSR-TR-76-1113 NL

3 OF 7
ADA031017


1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

Hence the discrete filter step response according to (4.c.6) is

$$\frac{1}{1-z^{-1}} + \frac{\frac{a-b}{b}}{1-e^{-T/b}\,z^{-1}} \qquad (4.c.10)$$

D. Design by approximating derivative by difference.

The idea (discussed in Section 4.B resulting in (4.B.19)) of approximating a continuous derivative uniformly sampled at nT by the first backward difference divided by T can be used directly on the differential equati (4.1) in the following way.

Define:

$$\nabla^{(0)}[z(n)] = z(n) \qquad (4.D.1)$$

and

$$\nabla^{(1)}[z(n)] = \frac{1}{T}\Delta_b[z(n)] = \frac{z(n) - z(n-1)}{T} \qquad (4.D.2)$$

and

$$\nabla^{(k)}[z(n)] = \nabla^{(1)}\left[\nabla^{(k-1)}[z(n)]\right] \qquad (4.D.3)$$

Let $v(n) = u(nT)$ and $z(n) = y(nT)$.

For $\left.\frac{d^k y(t)}{dt^k}\right|_{t=nT}$ consider the approximation $\nabla^{(k)}[z(n$

Applying this approximation to (4.1), get

$$\sum_{k=0}^{n} C_k \nabla^{(k)}[z(n)] = \sum_{k=0}^{M} d_k \nabla^{(k)}[r(n)] \qquad (4.D.4.)$$

Recalling $\mathcal{Z}\left[\Delta_b^{(k)}[f(n)]\right] = (1-z^{-1})^k F(z)$

and taking the Z-transform of (4.D.4) we get

$$K(z) = \frac{Z(z)}{V(z)} = \frac{\sum_{k=0}^{M} d_k \left(\frac{1-z^{-1}}{T}\right)^k}{\sum_{k=0}^{N} c_k \left(\frac{1-z^{-1}}{T}\right)^k} \qquad (4.D.5)$$

Comparing (4.D.5) with (4.2) we see that

$$K(z) = H(s)\Big|_{s=\frac{1-z^{-1}}{T}} \qquad (4.D.6)$$

so that this technique does correspond to a mapping of the s-plane to the z-plane. To investigate stability preserving properties of this mapping, solve for z in terms of s,

$$\begin{aligned} z &= \frac{1}{1-sT} = \frac{1}{1-\sigma T - j\omega T} \\ &= \frac{1}{2}\left[1 + \frac{(1-\sigma T)+j\omega T}{(1-\sigma T - j\omega T}\right] \end{aligned} \qquad (4.D.7)$$

For $s=j\omega$, then (4D.7) becomes

$$\begin{aligned} z &= \frac{1}{2}\left[1 + \frac{1+j\omega T}{1-j\omega T}\right] \\ &= \frac{1}{2}\left[1 + e^{j2\tan^{-1}\omega T}\right] \end{aligned} \qquad (4.D.8)$$

so the $j\omega$ axis maps into the circumference of a circle in the z-plane of radius 1/2 centered at z=1/2. For s in the LHP, then in (4.D.7) we have $\sigma<0$ so,

$$z = \frac{(1-\sigma T)+j\omega T}{(1-\sigma T)^2+(\omega T)^2} = Re[z] + jIm[z] \tag{4.D.9}$$

so that the LHP maps into the inside of the above circle. Thus, stability is preserved, but $j\omega$ does not map to the circumference of the unit circle in the z-plane.

As mentioned before, as T is made smaller and smaller we expect the approximation to get better and better. This is borne out by (4.D.7) since for $s=j\omega$ and for small T the spectrum of the discrete filter is concentrated on the circle mentioned following (4.D.8) near z=1. But such small T goes against the computation time needed in the filter implementation.

E. Other possible time domain designs.

Several other possible time domain design techniques will be mentioned for completeness.

1. Approximating continuous integration by a discrete integration is a basic idea which stems from an existing analog network (perhaps canonic) employing only integrators, multipliers, and summers (see Section 1). Starting with this analog network, each integrator can be replaced by a numerical approximation over each sample period T. In general the numerical approximation involves approximating the continuous integrand by a finite degree polynomial over each sampling period T. The zero degree polynomial approximation is the familar rectangular integration method, the first degree polynomial approximation is the trapezoidal integration method, etc. Each of these could be set up to provide an over-estimate or an under-estimate.

   Other numerical techniques exist, e.g., predictor-corrector, Range-Kutta, etc., and detailed discussions can be found in any good text on numerical methods such as Hildebrand [5].

2. Matching state equation solutions at sampling instants is a reasonable idea. Starting from an existing single input continuous system described by state equations of the form

$$\frac{d}{dt}\underline{x}(t) = \underline{F}\,\underline{x}(t) + \underline{G}\,u(t) \tag{4.E.1}$$

$$y(t) = \underline{C}\,\underline{x}(t) + d\,u(t) \tag{4.E.2}$$

then if input u(t) is such that it can be represented by a piecewise constant function

$$\hat{u}(t) = u(kT), \qquad kT \leq t < (k+1)T \tag{4.E.3}$$

it is possible to determine the output y(t) at the discrete times t=o,T, 2T, ... by means of the discrete state equations

$$\underline{x}((k+1)T) = \underline{A}(T)\,\underline{x}(kT) + \underline{B}(T)\,u(kt) \tag{4.E.4}$$

$$y(kT) = \underline{C}\,\underline{x}(kT) + d\,u(kT). \tag{4.E.5}$$

where

$$\underline{A}(T) = e^{ET} \quad , \quad \underline{B}(T) = \int_0^T e^{\underline{F}\lambda}\,d\lambda\,\underline{G} \tag{4.E.6}$$

the matrices $\underline{A}$ (T) and $\underline{B}$(T) have to be evaluated only once for a given sampling period T. The evaluation is usually done by some type of iterative procedure.

There are only two sources of error in this technique: 1) that coming from the above iterative procedure, and 2) the approximation of u(t) by $\hat{u}$(t).

Now from (4.E.4) and (4.E.5), network realizations using unit delays, multipliers, and summers (see Section 1) can be obtained.

FREQUENCY DOMAIN TECHNIQUES

F. Fast Fourier Transform method of nonrecursive filtering.

The nonrecursive filter described by (4.A.1) can be implemented using transform techniques in the following way. The discrete Fourier transform (DFT) of the sampled input is, for a total of N sample points,

$$\mathcal{F}_d\left[u(n)\right] = U(k) = \sum_{n=0}^{N-1} u(n)\,e^{-j\frac{2\pi kn}{N}} \quad , k=0,1,\cdots,N-1 \tag{4.F.1}$$

and the inverse discrete Fourier transform (IDFT) is

$$\mathcal{F}_d^{-1}[U(k)] = u(n) = \frac{1}{N}\sum_{k=0}^{N-1} U(k)\, e^{\frac{j2\pi kn}{N}}, \quad n=0,1,\cdots N-1 \tag{4.F.2}$$

Similarly, the DFT of the sampled impulse response is

$$\mathcal{F}_d[h(n)] = H(k) \tag{4.F.3}$$

then, from transform theory, the DFT of the sampled output is

$$\mathcal{F}_d[y(n)] = Y(k) = H(k)\,U(k) \tag{4.F.4}$$

and

$$y(n) = \mathcal{F}_d^{-1}[Y(k)] \tag{4.F.5}$$

Clearly, this transform procedure is an alternate way of doing the convolution (4.A.1). At first glance it appears to be a very inefficient method since (4.F.1) and (4.F.3) must be calculated, these multiplied as (4.F.4), and then (4.F.5) calculated. However, when N is large, say greater than fifty, and is of the form of integer two raised to an integer power, then the fast fourier transform technique makes this transform procedure computational advantageous over the direct calculation of the convolution (4.A.1). See, for example, reference [2].

G. Design by bilinear transformation (Tustin method)

We observed that the method of impulse response matching in Section 4.B had the problem of aliasing. This problem could be avoided if a transformation can be found that will map the entire s-plane into the horizontal strip in the s_1-plane bounded by the lines $s_1 = -j\frac{\omega_s}{2}$ and $s_1 = j\frac{\omega_s}{2}$ and which will periodically repeat this map in each of the other horizontal strips bounded by the lines $s_1 = j(n-\frac{1}{2})\omega s$ and $s_1 = j(n+\frac{1}{2})\omega s$ where n is an integer, and then Section 4.B applied between the s_1-plane and the z-plane.

A transformation with these properties is

$$s = \frac{2}{T} \tanh \frac{s_1 T}{2} \qquad (4.G.1)$$

Now substituting $z^{-1} = e^{-s_1 T}$

we get

$$s = \frac{2}{T} \, \frac{e^{\frac{s_1 T}{2}} - e^{-\frac{s_1 T}{2}}}{e^{\frac{s_1 T}{2}} + e^{-\frac{s_1 T}{2}}} = \frac{2}{T} \, \frac{1 - e^{s_1 T}}{1 + e^{s_1 T}}$$

$$= \frac{2}{T} \, \frac{1 - z^{-1}}{1 + z^{-1}} \qquad (4.G.2)$$

This transformation is related to numerical integration (see Section 4.E) by the trapezoidal rule in the following way. Consider the continuous system defined by the differential equation

$$d_1 \frac{d}{dt} y(t) + d_0 y(t) = c_0 u(t)$$

$$\text{now} \quad y(t) = \int_{t_0}^{t} \frac{d}{dt_1} y(t_1) dt_1 + y(t_0) \qquad (4.G.3)$$

so if $t = nt$ and $t_0 = (n-1)T$, then

$$y(nT) = \int_{(n-1)T}^{nT} \frac{d}{dt_1} y(t_1) dt_1 + y((n-1)T)$$

Approximating this integral by the trapezoidal rule, then

$$y(nT) = y((n-1)T) + \frac{T}{2} \left[\frac{dy(t)}{dt}\bigg|_{t=nT} + \frac{dy(t)}{dt}\bigg|_{t=(n-1)T} \right] \qquad (4.G.4)$$

but from (4.G.3)

$$\frac{dy(t)}{dt}\bigg|_{t=nT} = \frac{-d_0}{d_1} y(nT) + \frac{c_0}{d_1} u(nT)$$

and using this in (4.G.4) we get

$$\left[z(n) - z(n-1) \right] = \frac{T}{2} \left[\frac{-d_0}{d_1} \left(z(n) + z(n-1) \right) + \frac{c_0}{d_1} \left(v(n) + v(n-1) \right) \right]$$

where $z(n) = y(nT)$ and $v(n) = u(nT)$

Taking the Z-transform of the last equation we get

$$K(z) = \frac{Z(z)}{V(z)} = \frac{c_0}{d_1 \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}} + d_0} \tag{4.G.5}$$

Taking the Laplace Transform of (4.G.3) we get

$$H(s) = \frac{Y(s)}{U(s)} = \frac{c_0}{d_1 s + d_0} \tag{4.G.6}$$

so comparing the last two equations we see that

$$K(z) = H(s)\Big|_{s = \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}}} \tag{4.G.7}$$

The above equation can be shown to hold in general.

Consider (4.G.2) when $z=e^{j\theta}$,
then

$$s = \frac{2}{T} \frac{1-e^{-j\theta}}{1+e^{-j\theta}} = \frac{2}{T} j \tan\frac{\theta}{2} = \sigma + j\omega$$

$$\text{so } \sigma = 0 \text{ and } \frac{T\omega}{2} = \tan\frac{\theta}{2}$$

This is shown in figure 4.G.1

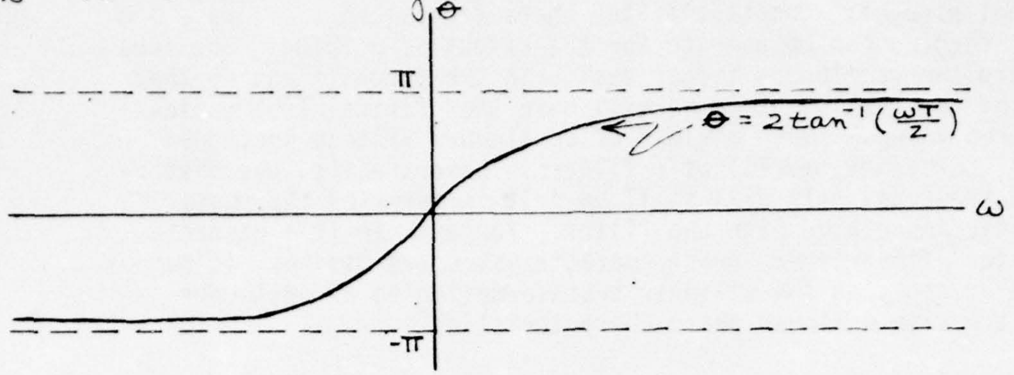


Figure 4.G.1

from which it is clear that the positive and negative imaginary axis of the s-plane are mapped into the upper and lower halves of the unit circle in the z-plane. Also inverting (4.G.2) and checking $|z|$ for $Re[s]<0$ shows that the LHP goes into the unit circle, thus preserving stability.

So, the bilinear transformation eliminates the aliasing problem and preserves stability, but introduces a distortion (warping) in the frequency axis, indicated by Figure 4.G.1. Hence this method is useful only when this distortion can be tolerated or compensated for (i.e. by prewarping.)

This frequency warping is best seen between the s-plane and the s_1-plane from (4.G.1). Let $s_1=j\omega_1$, and $s=j\omega$ in this equation and we get

$$\frac{\omega T}{2} = \tan\frac{\omega_1 T}{2} \qquad (4.G.8)$$

and the deviation from linearity of this equation is shown in fig. 4.G.2.

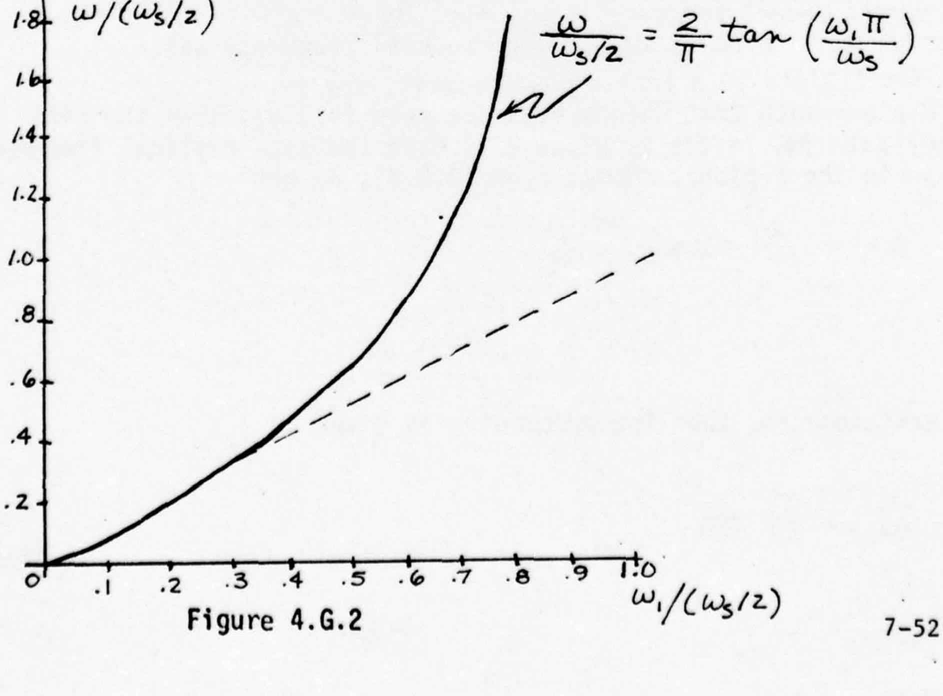


Figure 4.G.2

Now if the continuous filter is such that its magnitude characteristics approximate an ideal piecewise-constant filter characteristics, i.e., pass and stop bands, then we can compensate for the effect of warping. The idea is to prewarp the continuous filter design in the opposite way so that then applying the tustin method we will have the critical frequencies at the desired values. Such a class of continuous filters includes Butterworth, Chebyshev, and elliptic filters. Nevertheless, the distortion in the frequency axis will still be felt in terms of the phase characteristic associated with the filter. For example if a discrete lowpass filter with a linear phase characteristic was desired, it cannot be obtained by applying the bilinear transformation to a continuous lowpass filter with a linear phase characteristic.

Also if the bilinear transform is applied to a continuous filter whose magnitude characteristics are not essentially piecewise constant over the major part of the Nyquist frequency interval, then the frequency warping can yield an unsatisfactory discrete filter. For example, consider a wideband continuous differentiating filter described by H(s) = **s**. For this, figure 4.G.2 can be interpreted as the magnitude frequency response of the resulting discrete filter.

A computer program to incorporate the Tustin design method is referred to Reference [6].

EXAMPLE. The Tustin method has been used by Hughes Aircraft Company in their design of the digital autopilot PDAP. Consider the prototype continuous filter to be a lead-lag filter described by

$$H(s) = \frac{\tau_1 s+1}{\tau_2 s+1} \qquad (4.G.9)$$

The time constants τ_i in this transfer function are the reciprocal of the critical (corner) frequencies ω_i, i.e., $\tau_i = \frac{1}{\omega_i}$
Now what we want to do is to prewarp the real frequency axis, (jw) of the s-plane to a real frequency axis, say jw, in the $\hat{s}$-plane such that when $\hat{\omega}$ replaces ω in (4.G.8), then the real frequency axis $j\omega_1$ in the s_1 plane will have the same critical frequencies as did $j\omega$ in the s-plane. Hence from (4.G.8), we get

$$\omega_1 = \frac{2}{T} \tan^{-1} \hat{\omega} \frac{T}{2}$$

and to achieve $\omega_1 = \omega$, then the prewarping is given by

$$\hat{\omega} = \frac{2}{T} \tan \frac{\omega T}{2} \qquad (4.G.10)$$

Under (4.G.10) the prewarped time constants are

$$\hat{\tau}_i = \frac{1}{\hat{\omega}_i} = \frac{1}{\frac{2}{T}\tan\frac{\omega_i T}{2}} = \frac{\frac{T}{2}}{\tan\frac{T}{2\tau_i}} = \frac{T}{2}\tau_{\omega_i} \qquad \textbf{(4.G.11)}$$

where $\tau_{\omega_i} = \left(\tan\frac{T}{2\tau_i}\right)^{-1}$

Now the prewarped prototype continuous filter is

$$H(\hat{s}) = \frac{\hat{\tau}_1\hat{s}+1}{\hat{\tau}_2\hat{s}+1} = \frac{\frac{T}{2}\tau_{\omega_1}\hat{s}+1}{\frac{T}{2}\tau_{\omega_2}\hat{s}+1} \qquad \textbf{(4.G.12)}$$

Applying the Tustin method to this equation, get

$$K(z) = H(\hat{s})\Big|_{\hat{s}=\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}}$$

$$= \frac{\tau_{\omega_1}\left(\frac{1-z^{-1}}{1+z^{-1}}\right)+1}{\tau_{\omega_2}\left(\frac{1-z^{-1}}{1+z^{-1}}\right)+1} = \frac{\tau_{\omega_1}(1-z^{-1})+(1+z^{-1})}{\tau_{\omega_2}(1-z^{-1})+(1+z^{-1})}$$

$$= \frac{(1+\tau_{\omega_1})+(1-\tau_{\omega_1})z^{-1}}{(1+\tau_{\omega_2})+(1-\tau_{\omega_2})z^{-1}}$$

$$= \frac{1+\tau_{\omega_1}}{1+\tau_{\omega_2}} \quad \frac{1+\frac{1-\tau_{\omega_1}}{1+\tau_{\omega_1}}z^{-1}}{1+\frac{1-\tau_{\omega_2}}{1+\tau_{\omega_2}}z^{-1}}$$

$$= K_1 \quad \frac{1+C_1 z^{-1}}{1+D_1 z^{-1}}$$

(4.G.13)

which is the Hughes result. This result was independently worked out in a different format by Capt. Moriarty in DLMA.

H. Some other possible frequency domain methods

Many other frequency domain design methods have been advanced. Several of the more qppealing ones will be briefly described.

1. The "matched Z transformation" is an extension of the Tustin method in that both real and imaginary parts of poles and zeros are warped, compared to just the imaginary part as in the Tustin method. The idea is that this can give some control over the phase characteristics. This type of prewarping is illustrated by considering the simple prototype continuous filter described by

$$H(s) = \frac{1}{s+d_o}$$

Making the follwoing replacements

$$s \longrightarrow \frac{2}{T}\tanh\left(\frac{sT}{2}\right)$$

$$d_o \longrightarrow \frac{2}{T}\tanh\left(\frac{d_oT}{2}\right)$$

we get

$$K(z) = \left[\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}} + \frac{2}{T}\frac{1-e^{-d_oT}}{1+e^{-d_oT}}\right]^{-1}$$

$$= \frac{T}{4}\frac{(1+e^{-d_oT})(1+z^{-1})}{1-e^{-d_oT}z^{-1}}.$$

See Reference [7].

2. Design based on minimization of mean square error at discrete frequencies.

Assume the discrete filter transfer function is

$$K(z) = A\prod_{k=1}^{N} \frac{1+a_k z^{-1}+b_k z^{-2}}{1+c_k z^{-1}+d_k z^{-2}} = AK_1(z)$$

and suppose the desired frequency response

$$K_d(z)\Big|_{z=e^{j\omega}}$$

is prescribed at the frequencies

$$\left\{\omega_i\right\}_{i=1}^{N}$$

Define the mean-squared error at these frequencies by

$$E = \sum_{i=1}^{N}\left[\left|K(e^{j\omega_i})\right| - \left|K_d(e^{j\omega_i})\right|\right]^2$$

Then it is possible to minimize E as a function of the parameters $(a_1,b_1,c_1,d_1 \ldots a_N,b_N,c_N,d_N,A)$. See References [8],[9]. The above procedure has been generalized to minimizing a weighted average of the error raised to the pth power [10]. A related procedure in the time domain has also been considered. [11].

3. Design of Nonrecursive Filters using "Windows"

A simple approach to obtaining a nonrecursive filter is to start with an infinite duration unit sample response $k_r(n)$ and to truncate it to a finite duration unit sample response $k_{nr}(n)$. This can be viewed as a product in the time domain of $k_r(n)$ and a finite duration "window" w(n), i.e.

$$k_{nr}(n) = kr(n)\, w(n)$$

This multiplication in the time domain corresponds to convolution in the frequency domain, so if w(n) is chosen such that the frequency content of its transform is concentrated mostly in a central lobe, say $|\omega| < \omega_o$ then the convolution will result in a close reproduction of the frequency response of kr(n). See References [4],[12].

REFERENCES

1. C. Jordan, Calculus of Finite Differences, Chelsea, 1960.

2. A. Oppenheim and R. Schafer, Digital Signal Processing, Prentice-Hall, 1975.

3. M. A. Soderstrand and J. K. Maulden, Digital Filter Design Manual, Sandia Corp. SCL-DR-71-46, January 1972.

4. F. F. Kno and J. F. Kaiser, System Analysis by Digital Computer, Wiley, 1967.

5. F. B. Hildebrand, Introduction to Numerical Analysis, McGraw-Hill, 1956.

6. Themis Research Staff, Report ECOM-69-0365-F, Rensselaer Polytechnic Institute, Troy, New York, 1L181, pp.85-88.

7. B. Gold and C.M. Rader, Digital Processing of Signals, McGraw-Hill, 1969.

8. K. Steiglitz, Computer-Aided Design of Recursive Digital Filters, IEEE Trans on Audio and Electrcacoustics, Vol. AV-18,pp.123-129, June, 1970.

9. T. L. Walters, Optimal Design of Recursive Digital Filters, CSL Report R-628, September 1973, UILU-ENG 73-2230, U. of Illinois, Urbana, Illinois.

10. A. G. Deczky, Synthesis of Recursive Digital Filters using the minimum p Error Criterion, IEEE Trans. Audio Electroacoust., Vol. AU-20, pp. 257-263, October 1972.

11. C. S. Burrus and T. W. Parks, Time Domain Design of Recursive Digital Filters, IEEE Trans. Audio Electroacoust., Vol AU-18, No. 2, pp137-141, June 1970.

12. R. B. Blackman and J. W. Tukey, The Measurement of Power Spectra, Dover Publications, New York, 1958.

13. J. N.´Youngblood, AFATL TR in preparation

14. J. F. Delansky, Part I: A Common Review of Frequency Domain Theory for Continuous and Discrete Time Linear Time Invariant Systems, AFATL TR in preparation.

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT-PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

TRANSIENT THERMAL ANALYSIS OF EXTERNAL STORES

Prepared by: Dupree Maples Phd.

Academic Rank: Associate Professor

Department and University: Department of Mechanical Engineering
Louisiana State University

Assignment:

(Laboratory) Armament
(Division) Munitions
(Branch) Aircraft Compatibility

USAF Research Colleague: J. C. Key, Jr.

Date: August 22, 1975

Contract No.: F44620-75-C-0031

TRANSIENT THERMAL ANALYSIS OF EXTERNAL STORES

By

Dupree Maples Phd.

ABSTRACT

This report presents the results of a numerical method of evaluating the temperature distribution within externally carried bomb weapons subjected to aerodynamic heating. The specific weapon chosen to demonstrate the method is the Mark 84 bomb.

A general form of the heat conduction equation, solved by finite element, applicable to radial heat flow into bodies of revolution is presented. Application of this equation to various locations within the Mark 84 bomb is possible with the "TRAX" computer program.

Numerical results obtained on a CDC 6600 computing machine for a typical mission in which the bomb, initially at 100^{O}F, is subjected to flight at various Mach numbers and altitudes are included. The bomb was thermally analyzed for the cases of various distributions of the convective heat transfer coefficient applied to the external surface of the bomb. Analyses were conducted with and without the surface being insulated.

LIST OF FIGURES

NOMENCLATURE

T = Temperature at any point within the bomb

T_{AW} = Adiabatic wall temperature

t = Time

R_O = Outside radius of bomb

r = Element axis

z = Element axis perpendicular to r axis

k_r = Material conductivity in r direction

k_z = Material conductivity in z direction

C = Specific heat of material

ρ = Material density

h = Convective heat transfer coefficient

INTRODUCTION

High-velocity convection involves essentially two different phenomena:

1. Conversion of mechanical energy to thermal energy, resulting in temperature variations in the fluids;

2. Variation of the fluid properties as a result of the temperature variation.

Extremely high velocities in gases like air lead to very high temperatures, dissociation, mass concentration gradients, and thus mass diffusion, which further complicates the problem. However, we will restrict attention to the boundary layer with no chemical reaction; this means that for air, at least, we will not deal with temperatures greater than about 3500^{o}F, or Mach numbers greater than about 6.

Missile or aircraft flight at high speeds introduces new problems to the engineer. One of the more important problems that arises is caused by the skin temperatures (aerodynamic heating) that are attained at very high velocities. Adiabatic wall temperatures under these conditions can exceed temperature limitations of structural materials commonly used in the manufacture of such missiles. The payload of explosives also has a critical temperature level which may be exceeded.

The increase in flight speeds of modern aircraft has emphasized a great number of problems arising as a result of aerodynamic heating. Excessively high temperatures may not only affect the operational function of a weapon, such as chemical item for example, but in the case of explosives, cool-off may occur before the bomb is released. This is especially true for bombs which are carried externally by an aircraft, as by suspension under a wing. It is therefore important to have an understanding of heat transfer phenomena in bomb weapons.

This report documents some of the analytical investigation that has been done on aerodynamic heating of external carried weapons. The heating distributions on an external store (MK-84 nose section) have been determined for a typical flight mission.

OBJECTIVES AND SCOPE

The following research goals were outlined for the duration of the 12 week research program:

1. Convert the "TRAX" program developed by Arnold Engineering Development Center from an IBM 360 to a CDC 6600 computer.

2. Determine the effects of level changes in the convective heat transfer coefficient.

3. Investigate the effect of distribution changes in the convective heat transfer coefficient on the MK-84 bomb's temperatures.

4. Determine the amount of thermal insulation to control maximum regional temperatures.

5. To develop computer routines which will allow users to better display the results for quick analysis.

TECHNICAL APPROACH

This section of the report considers both the analytical model of the aerodynamic heating problem and the computer solution. The mathematical model is for transient heat conduction analysis of axisymetric bodies. The computer program is a finite element numerical solution to the mathematical model.

Mathematical Model

The transient heat conduction equation as presented by Kreith (1) for an axisymmetric body is

$$\frac{1}{r}\left[\frac{\partial}{\partial r}\left(k_r r\frac{\partial T}{\partial r}\right)\right] + \frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right) = \rho c \frac{\partial T}{\partial t}$$

The convective boundary condition is then

$$-k \left.\frac{\partial T}{\partial r}\right|_{r=R_O} = h \left(T - T_{AW}\right)_{r=R_O}$$

where T_{AW} is the adiabatic wall temperature and R_O is the outside radius of the MK-84 bomb. These equations were solved numerically to obtain temperature histories of the bomb.

Numerical Solution

There are two basic numerical techniques normally used to solve problems which have complex geometries and boundary conditions. These two methods are finite difference and finite element. The finite element method has proven to be advantageous when solving three-dimensional problems with irregular boundary conditions and nonhomogeneous properties as presented by Adams (2). Whereas the finite difference method approximates the governing differential equation, the finite element method approximates the solution. This approximation results from a minimization process based upon the theories of variational calculus.

The computer program TRAX developed by Rochelle (3) and obtained from AEDC should be competitive with finite difference programs as a general two-dimensional thermal analyzer. The TRAX program with some modernizations has been used in this study. A thermal analysis of the MK-84 nose section was performed.

Analyses were performed on the AFATL CDC 6600 computer. The computer code (TRAX) utilizes the finite-element technique to solve the transient heat conduction equation for axisymetric bodies. A complete description of this program is given in reference 3. The nose section without insulation in Figure 1 and with insulation in Figure 2, is an illustration of the finite element modeling required for the TRAX computer program. The thermophysical properties used are listed below:

Component	Material	Thermal Conductivity k(Btu/ft-sec-^{o}F)	Specific Heat C_p (Btu/lb-^{o}F)	Density ρ (Lb/ft^3)
Case	Steel	7.22×10^{-3}	0.120	491.0
Fuse	Telryl	4.583×10^{-5}	0.200	95.4
Explosive	Tritonal	7.388×10^{-5}	0.230	108.0
Insulation	AVCOAT 480	1.556×10^{-5}	0.290	40.0

The flight conditions for this investigation were based on the "typical mission profile" used by the Sandia personnel who performed a thermal analysis on a "Tactical bomb" (Reference 4). The altitude profile shows in Figure 3 and the Mach number profile illustrated in Figure 4 describe the typical mission.

With the properties and mission profile given above the next required input to the TRAX program is the boundary conditions. These are:

1. Initial temperature

   A temperature of 100^{o}F was selected to represent the bombs' temperature before flight.

2. Adiabatic wall temperature

   $(T_o = T\infty \left[1 + .2M\infty^2\right]$
   $T_{AW} = .97\ T_o)$
   Figure 5 presents the adiabatic wall temperatures calculated by the above equations and used in the computer analysis.

3. Convective heat transfer coefficient

   h, Btu/hr-ft^2-^{o}F

   Case I. These values are considered an approximate upper bound and is based on theoretical calculations given by Budenholzer (5) and shown in Figure 6.

   Case II. These values are considered an approximate median value and were obtained from Budenholzer (5) for the conditions of Mach number equal 2 and an altitude of zero. The values are shown in Figure 6.

   Case III. These values are obtained by subtracting 200 Btu/hr-ft^2-^{o}F from the values of Case II. Cases I, II, and III h is assumed to be independent of time.

Case IV. For this case the heat transfer coefficient was a function of time and position as illustrated in Figure 7. The values obtained closely match the typical mission described in Figures 3 and 4.

RESULTS

The following results were obtained in satisfying the objectives of this research program:

1. The "TRAX" program was successfully converted from an IBM 360 to a CDC 6600 computer. A test case was run on the IBM 360 at AEDC and the results for this case were duplicated on the CDC 6600.

2. The effects of level changes in the convective heat transfer coefficient are presented in Figures 8 through 11 for four nodes located seven inches from the nose of the MK-84 bomb. As presented in the figures, the higher the heat transfer coefficient the faster the bomb responds to a change in adiabatic wall temperature. This limits the time a given bomb can be flown at high speeds before damaging temperatures are obtained.

3. The effects of insulation was investigated by adding a 1/16 inch layer of insulation to the outside of the bomb case. The resulting reduction in temperatures at internal nodes is presented in Figures 12 through 15 for various heat transfer coefficient distributions. The surface temperature at a location of seven inches from the nose (as seen in the figures) closely follows the adiabatic wall temperature for the particular flight mission. In comparing Figure 12 with Figure 8, it can be seen that the maximum internal temperature is reduced from over 200^{o}F to under 130^{o}F as a result of adding the insulation.

4. Many other plots were made in analyzing the MK-84 bomb and given to Major Key for future reference. Due to the limited number of pages for this report, they were not included. The type of plots not included are:

   a. Variations of temperature with time for Cases I, II, III, and IV at cross sections of three and seventeen inches, with and without insulation.

   b. Bomb case temperature distributions subjected to heat transfer coefficients represented by Cases I, II, III, and IV.

   c. Surface temperature distributions at various times on a MK-84 bomb nose section subjected to heat transfer coefficients represented by Case I, II, III, and IV, with and without insulation.

CONCLUSIONS

A finite-element transient computer program was used to obtain temperature estimates for the nose section of a MK-84 bomb. The initial condition and mission profile simulated were similar to those used by the Sandia Corporation in their thermal analysis of bombs. The main conclusions of this investigation are:

1. The bombs' temperature is related to the convective heat transfer coefficient for the mission simulated.

2. Insulation added to the outside surface of the bomb greatly reduces the internal temperatures.

3. Insulation makes the bombs' internal nodes insensitive to changes in the convective heat transfer coefficients.

RECOMMENDATIONS

The following are recommended in the area of aerodynamic heating of weapons:

1. Compare the "TRAX" computer program with known analytical exact solutions.

2. Convert a three-dimensional transient heat conduction program to the CDC 6600 for aerodynamic heating analysis.

3. Develop a thermal laboratory so that the mathematical analysis and assumptions can be experimentally verified.

REFERENCES

1. Kreith, Frank, Heat Transfer. International Textbook Company, July 1968.

2. Adams, J. Alan and Rogers, David F., Computer-Aided Heat Transfer Analysis. McGraw-Hill Book Company, 1973.

3. Rochelle, James Kenneth, "TRAX - A Finite Element Program for Transient Heat Conduction Analysis of Axisymmetric Bodies" Master Thesis, University of Tennessee Space Institute, March 1973.

4. Private communications from Mr R. K. Matthews, Project Engineer, Aerodynamics Projects Branch, Von Karman Gas Dynamics Facility, ARO, Inc., June 1975.

5. Budenholzer, R. A., Goldsmith, A. and Nielsen, H. J., "Investigation of Problems of Temperature and Pressure Influencing the Design of Bomb Weapons," AFAC-TR-57-112, November 1957.

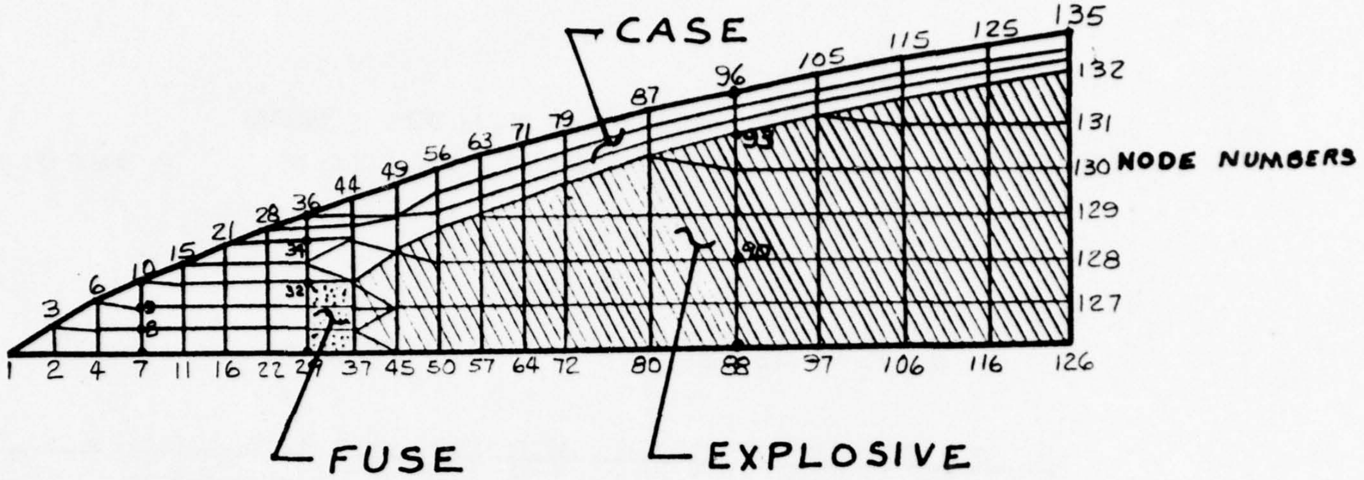


FIGURE 1. FINITE-ELEMENT MODEL OF MK-84 BOMB NOSE SECTION WITHOUT INSULATION

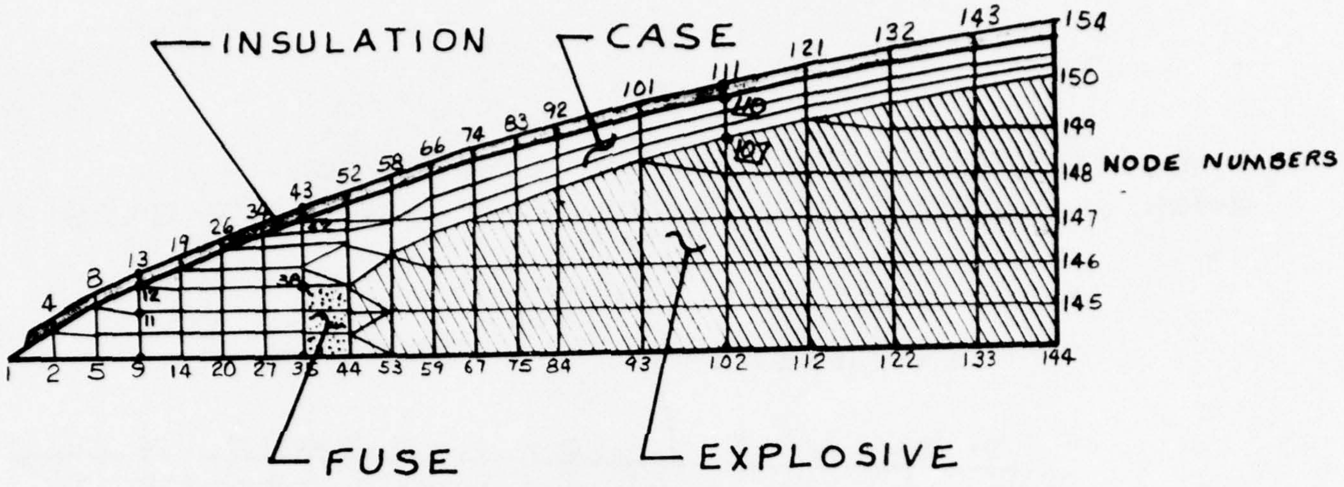


FIGURE 2. FINITE-ELEMENT MODEL OF MK-84 BOMB NOSE SECTION WITH INSULATION

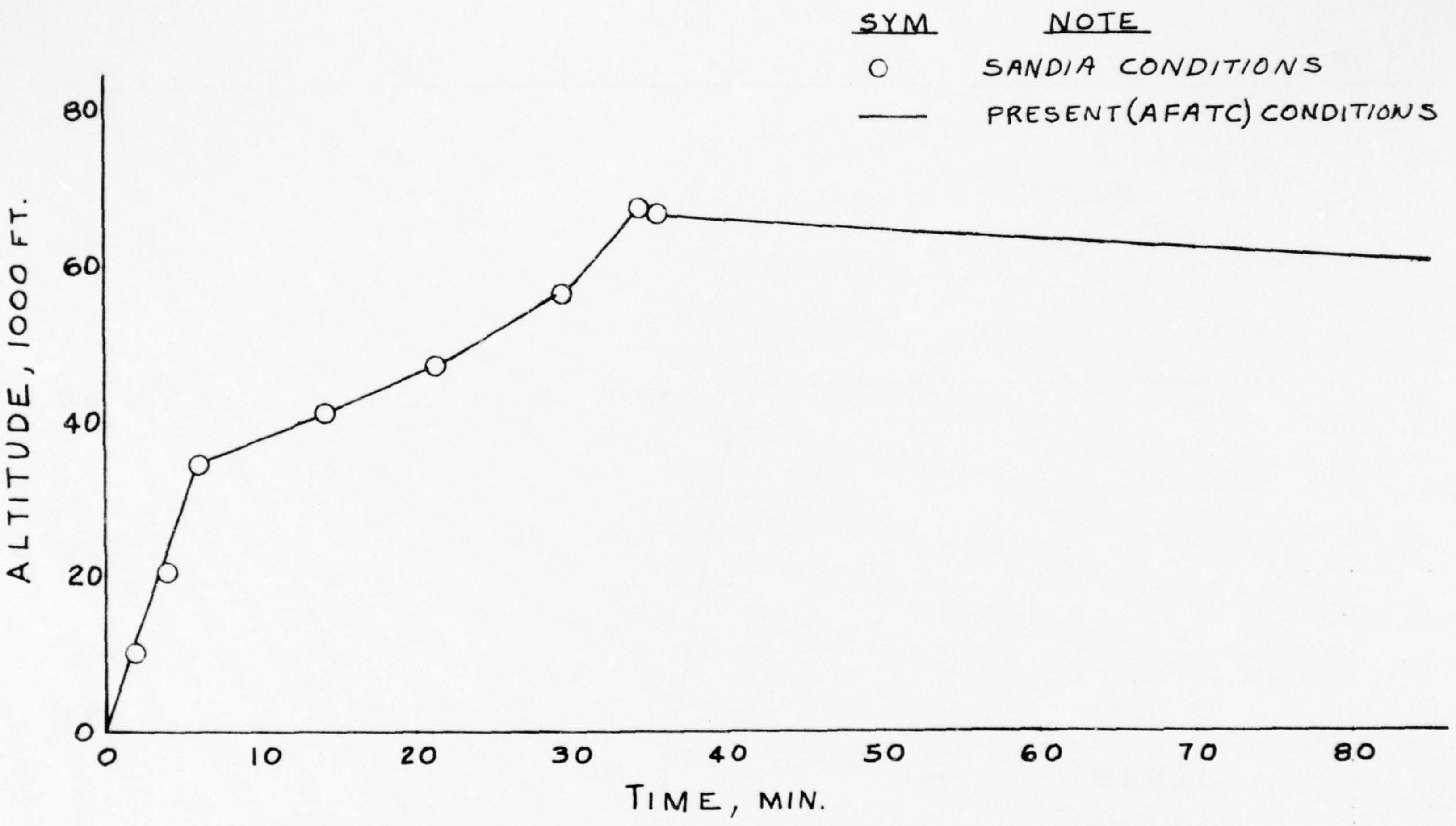


FIGURE 3. TYPICAL MISSON ALTITUDE PROFILE

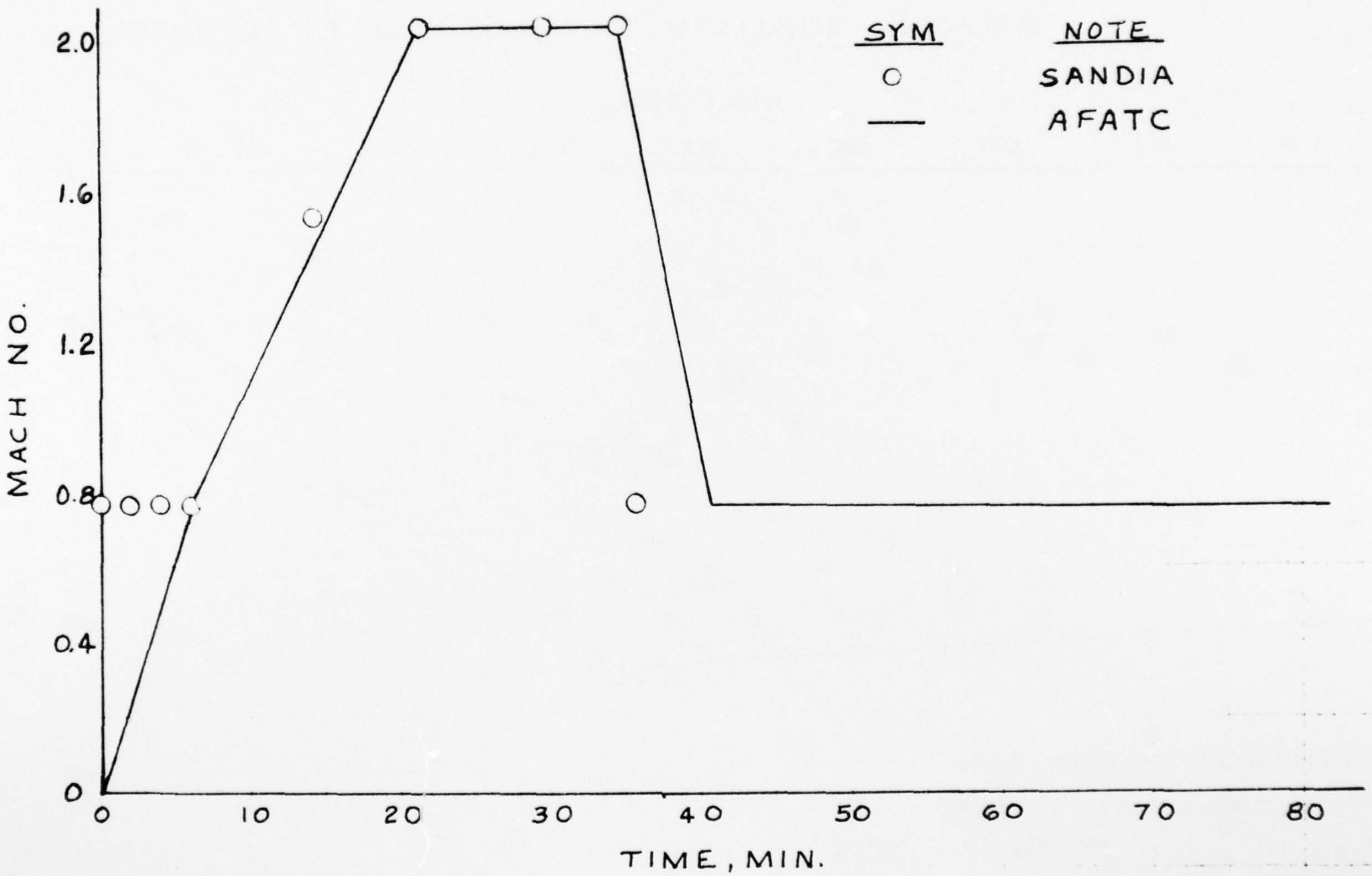


Figure 4. Typical Mission Mach No. Profile

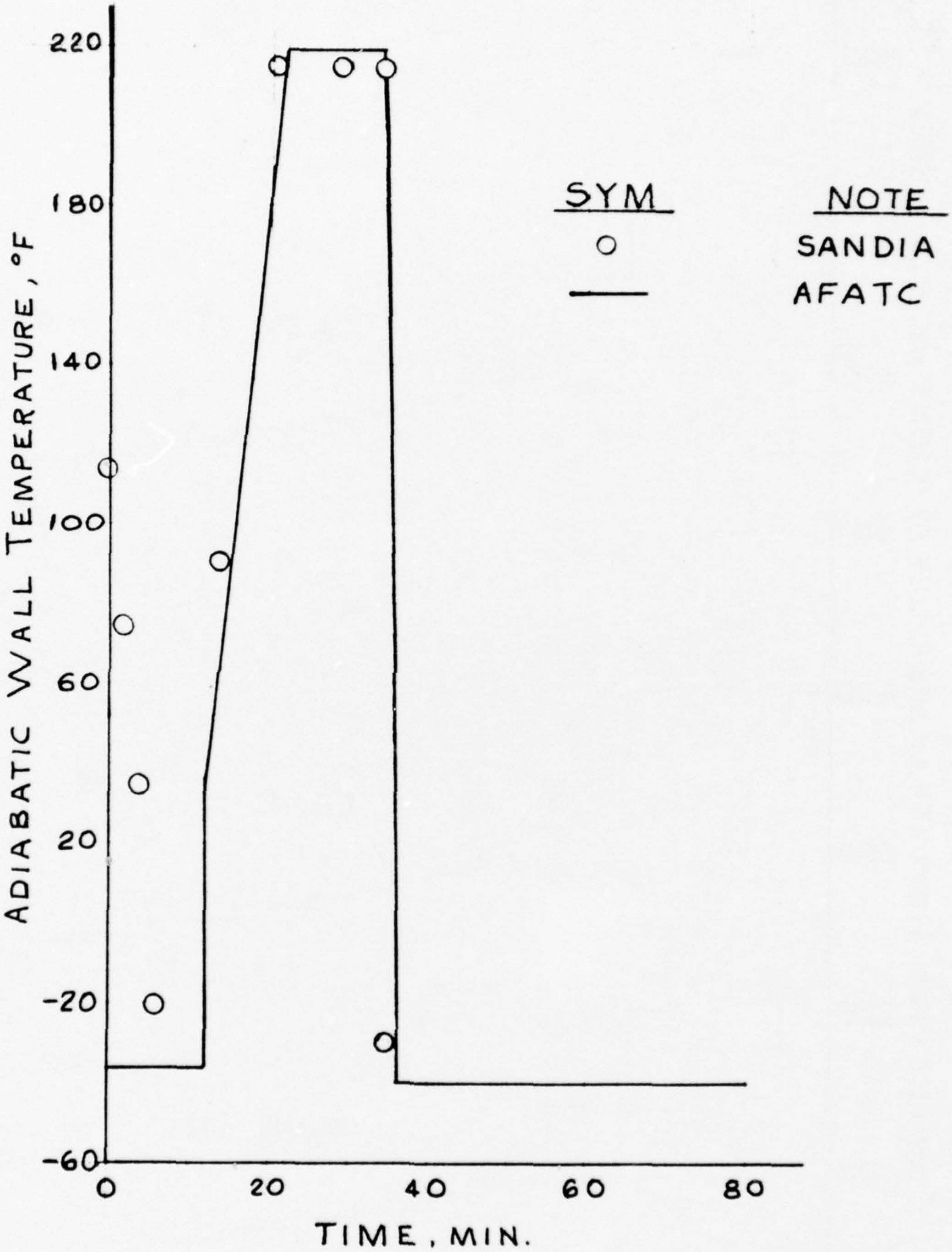


FIGURE 5. ADIABATIC WALL TEMPERATURE PROFILE FOR TYPICAL MISSON

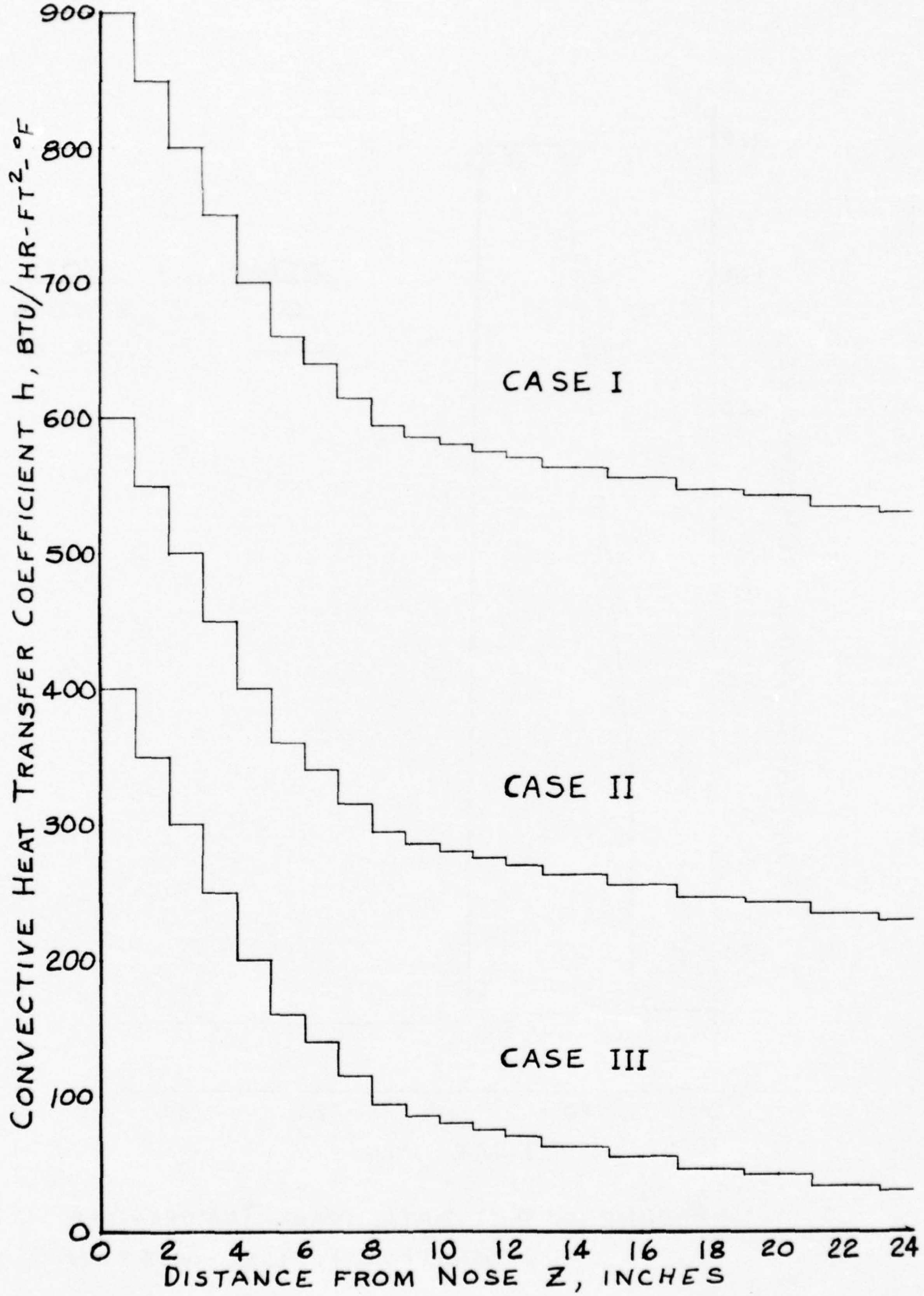


Figure 6. Distribution of the Heat Transfer Coefficient on the Outside Surface

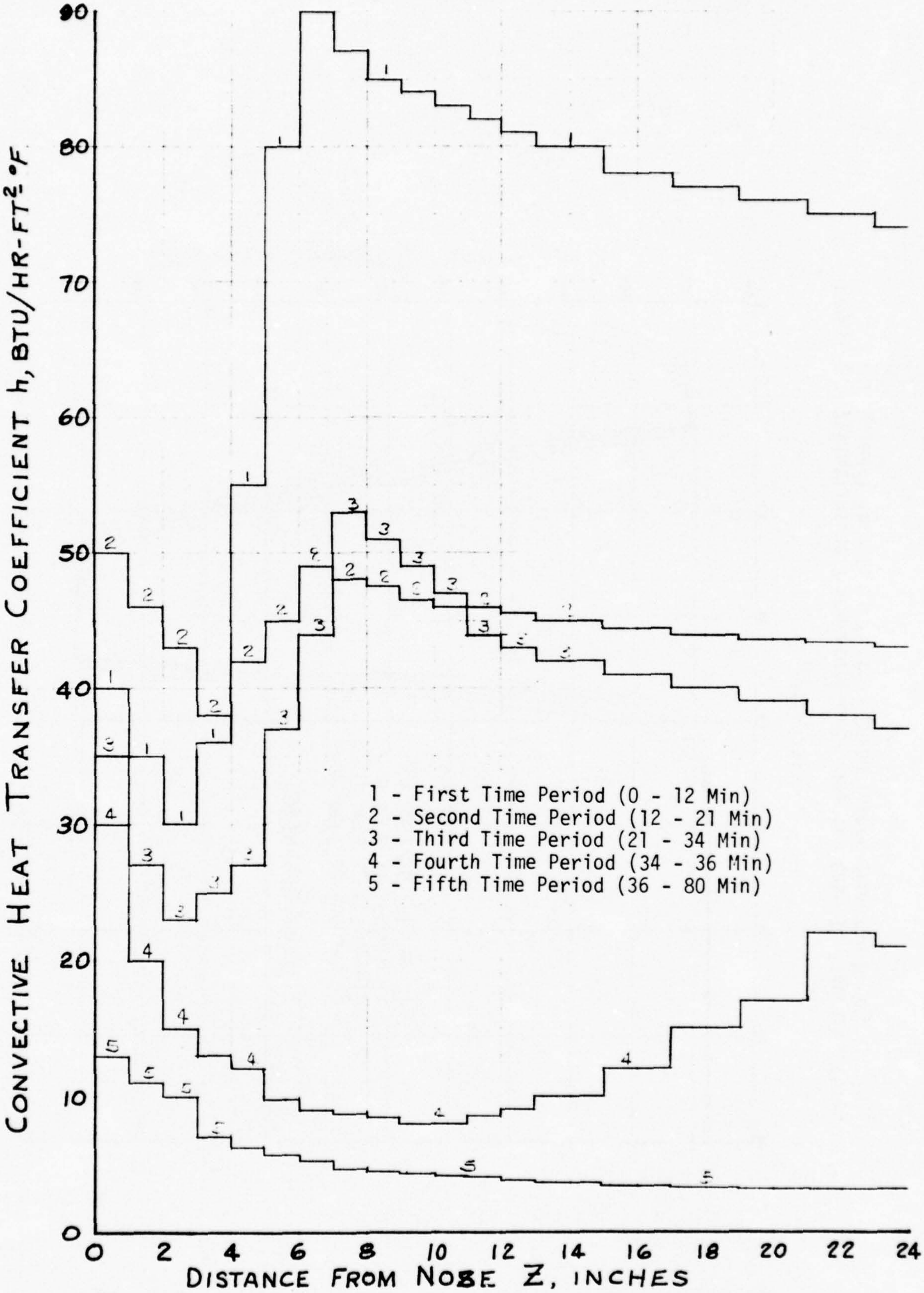


Figure 7. Variations of the Convective Heat Transfer Coefficient that Represents the Typical Flight Profile

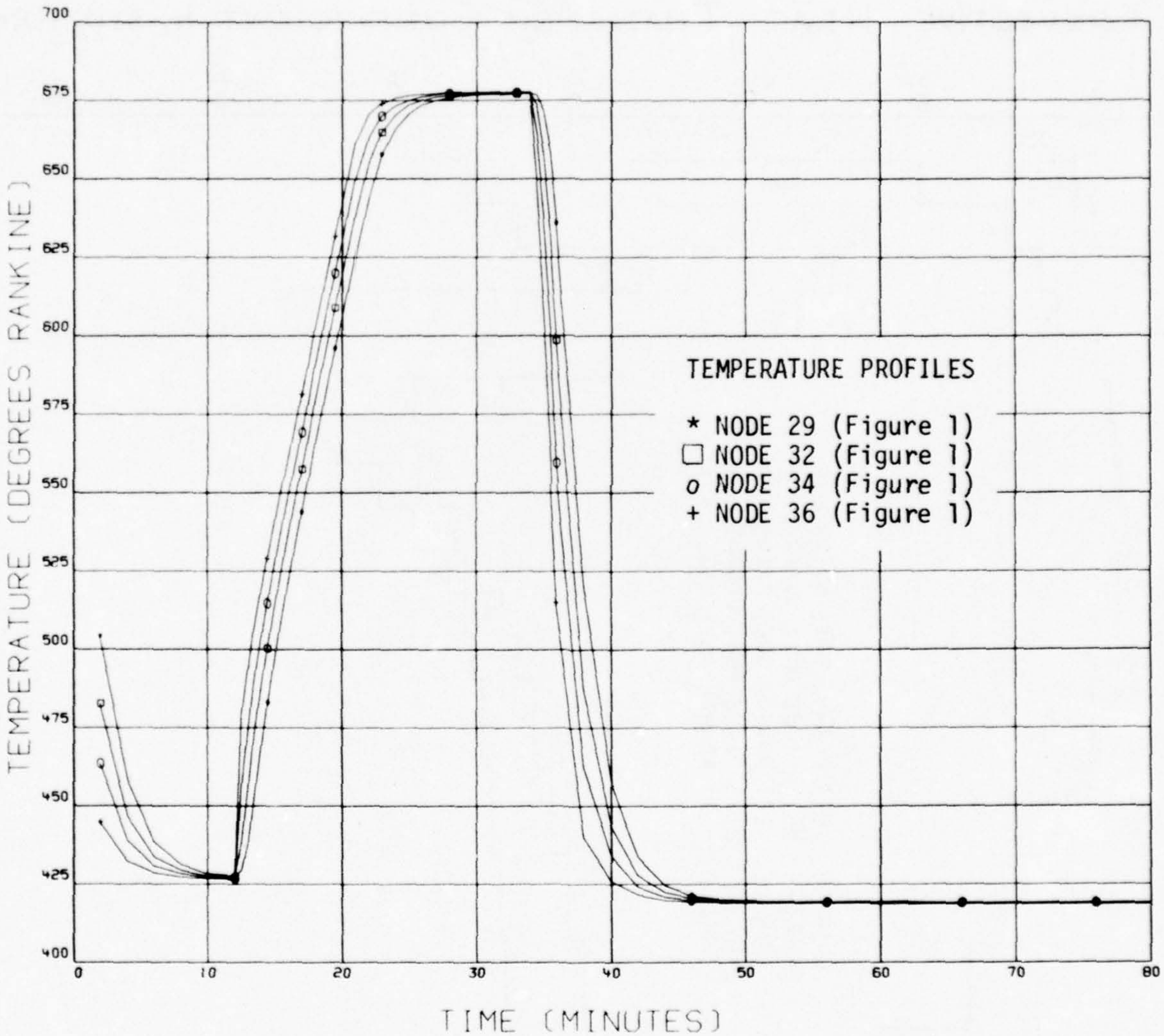


Figure 8. Variations of Temperature with Time for Case I, Figure 6, (High h Values) at a Cross Section Seven Inches from the Nose

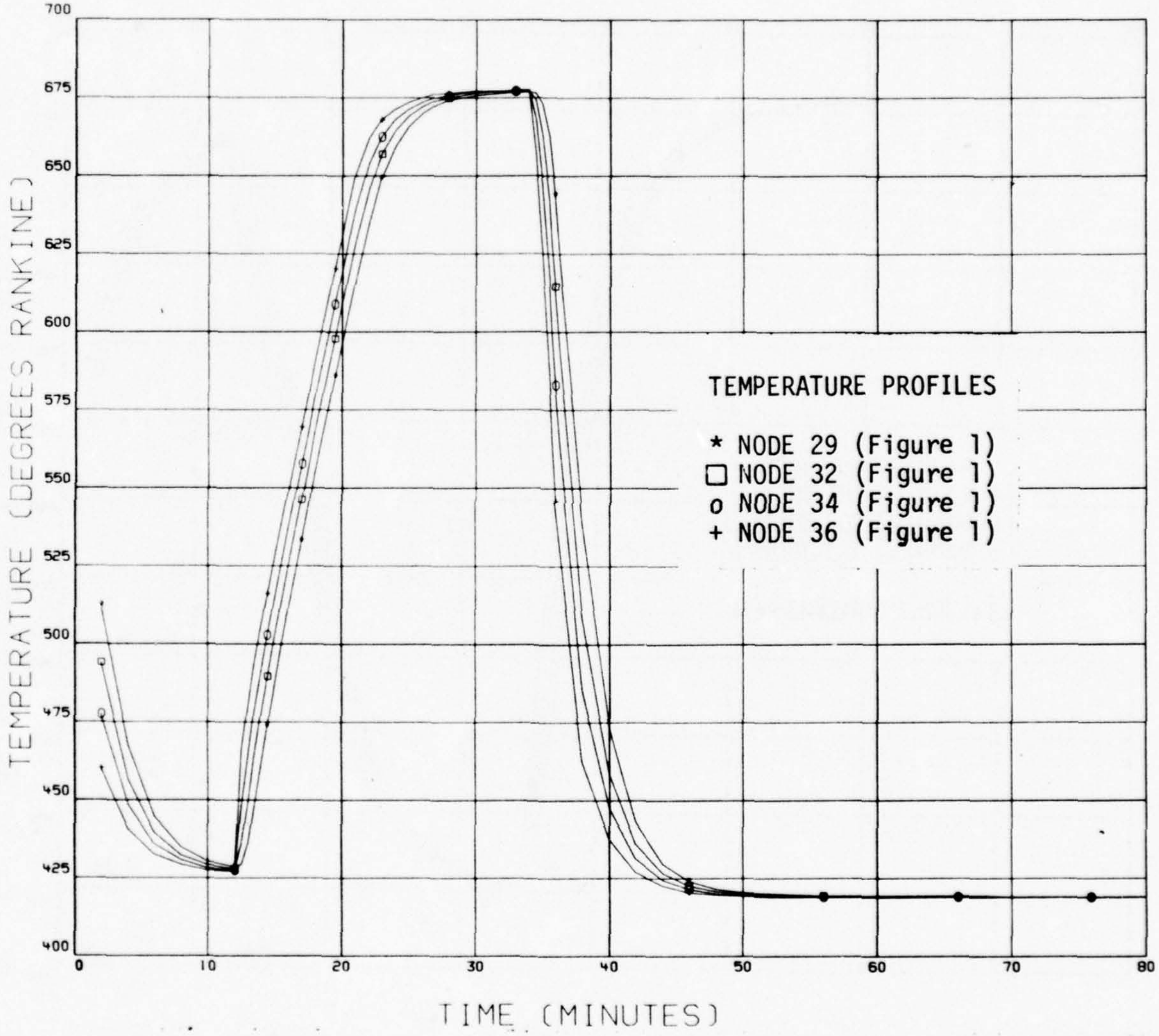


Figure 9. Variations of Temperature with Time for Case II, Figure 6, (Medium h Values) at a Cross Section Seven Inches from the Nose

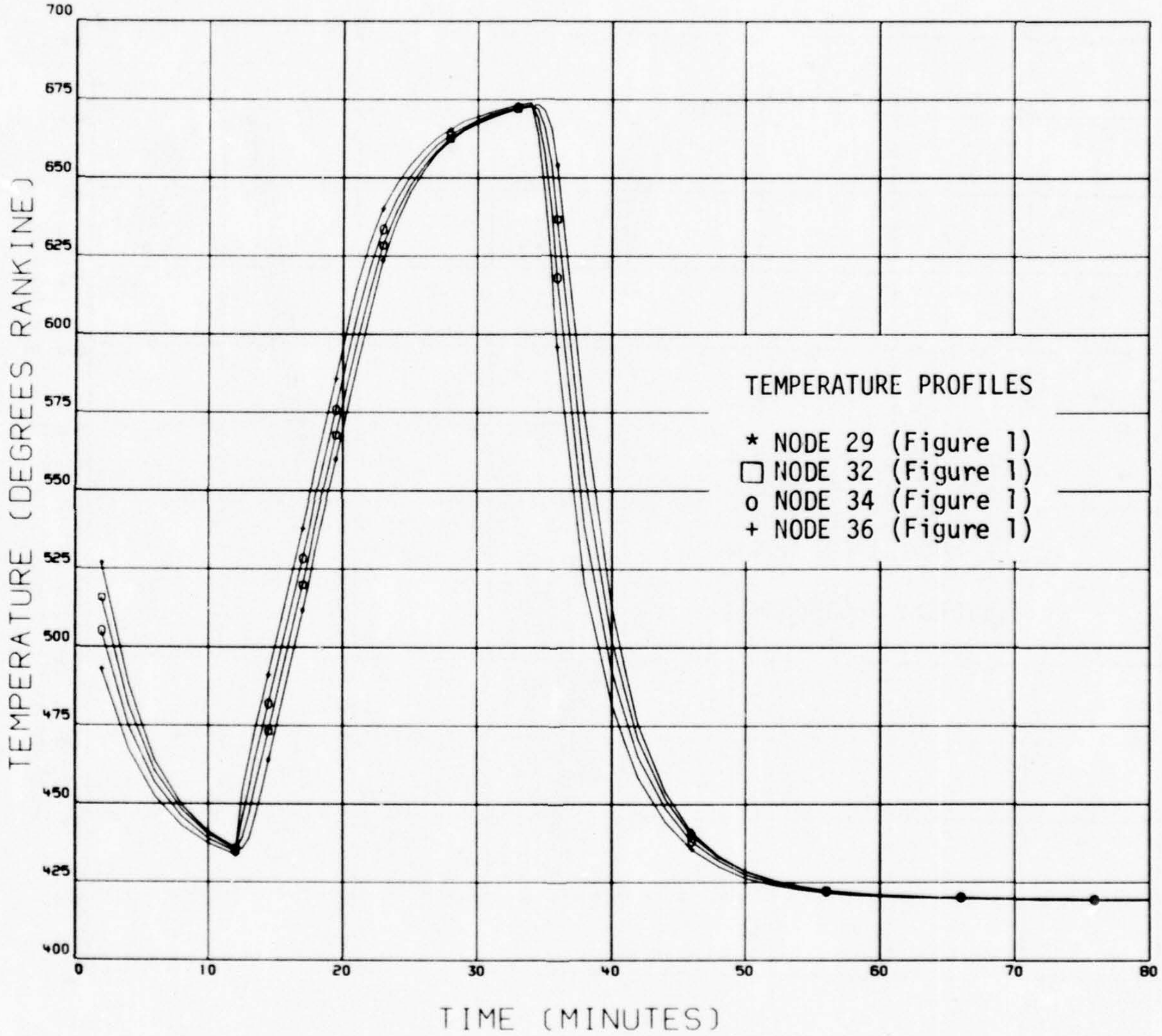


Figure 10. Variations of Temperature with Time for Case III, Figure 6, (Low h Values) at a Cross Section Seven Inches from the Nose

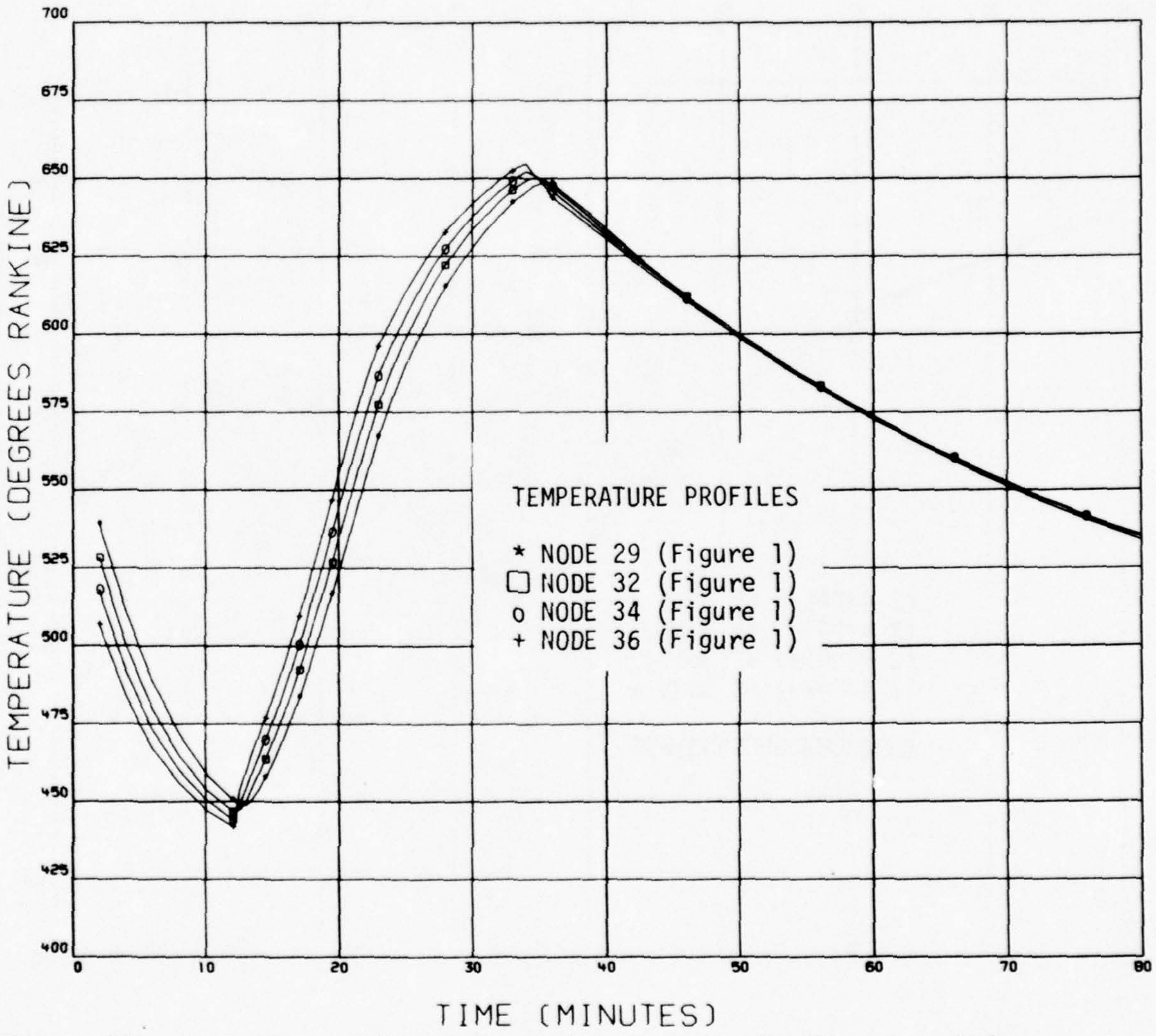


Figure 11. Variations of Temperature with Time for IV, Figure 7, [h(Pos. and Time)] at a Cross Section Seven Inches from the Nose

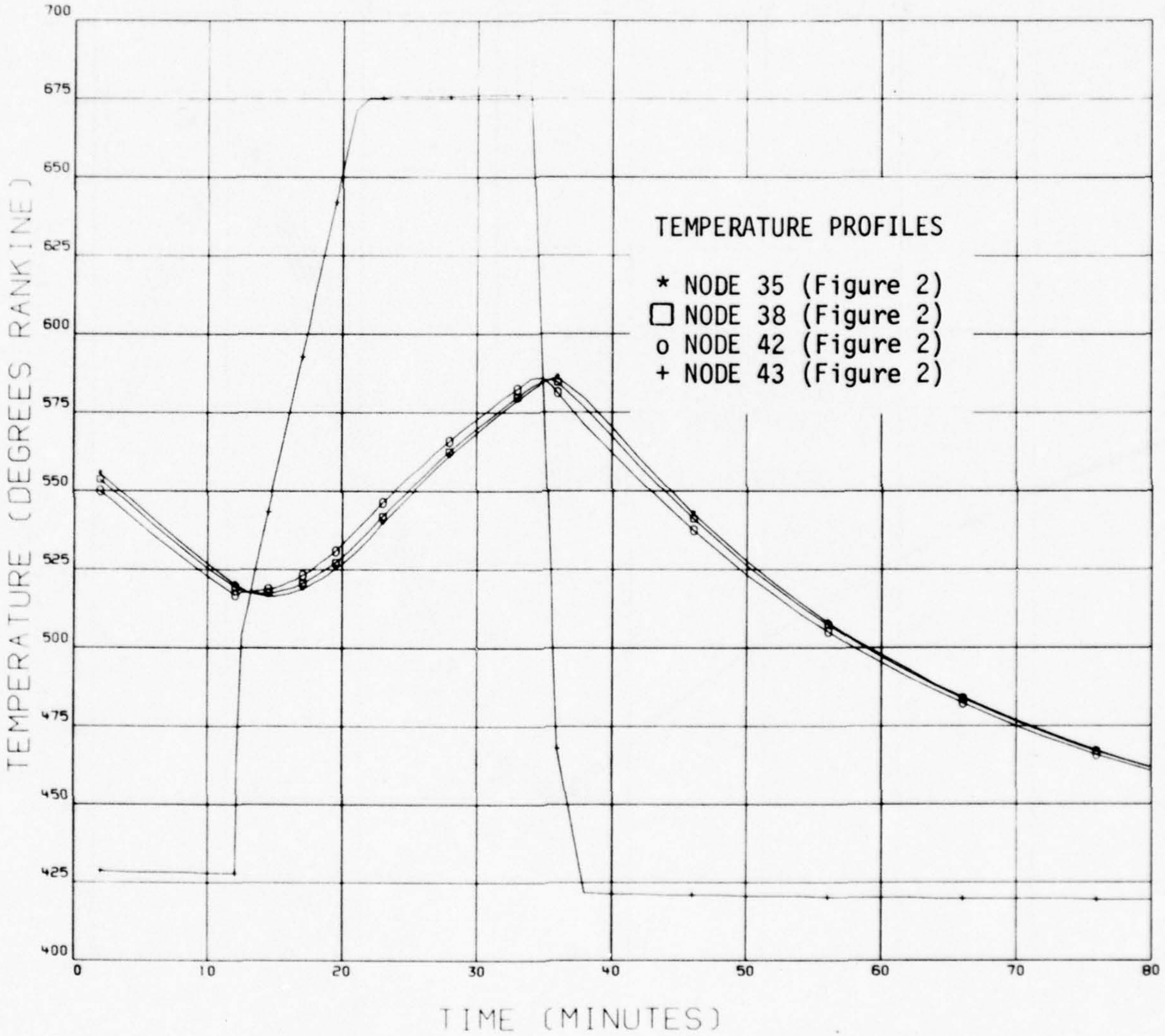


Figure 12. Variations of Temperature with Time for Case I, Figure 6, (with Insulation) at a Cross Section Seven Inches from the Nose

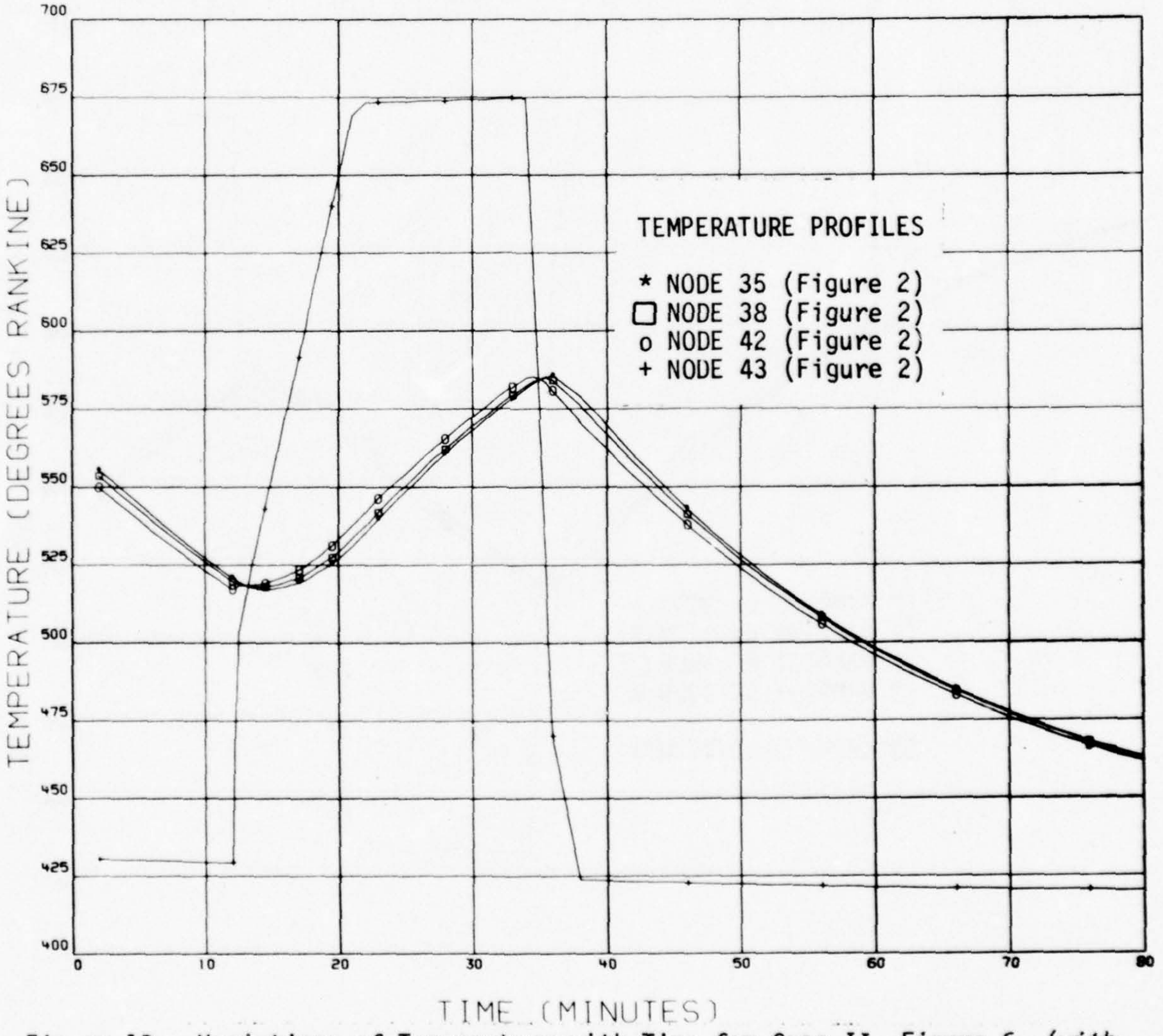


Figure 13. Variations of Temperature with Time for Case II, Figure 6, (with Insulation) at a Cross Section Seven Inches from the Nose

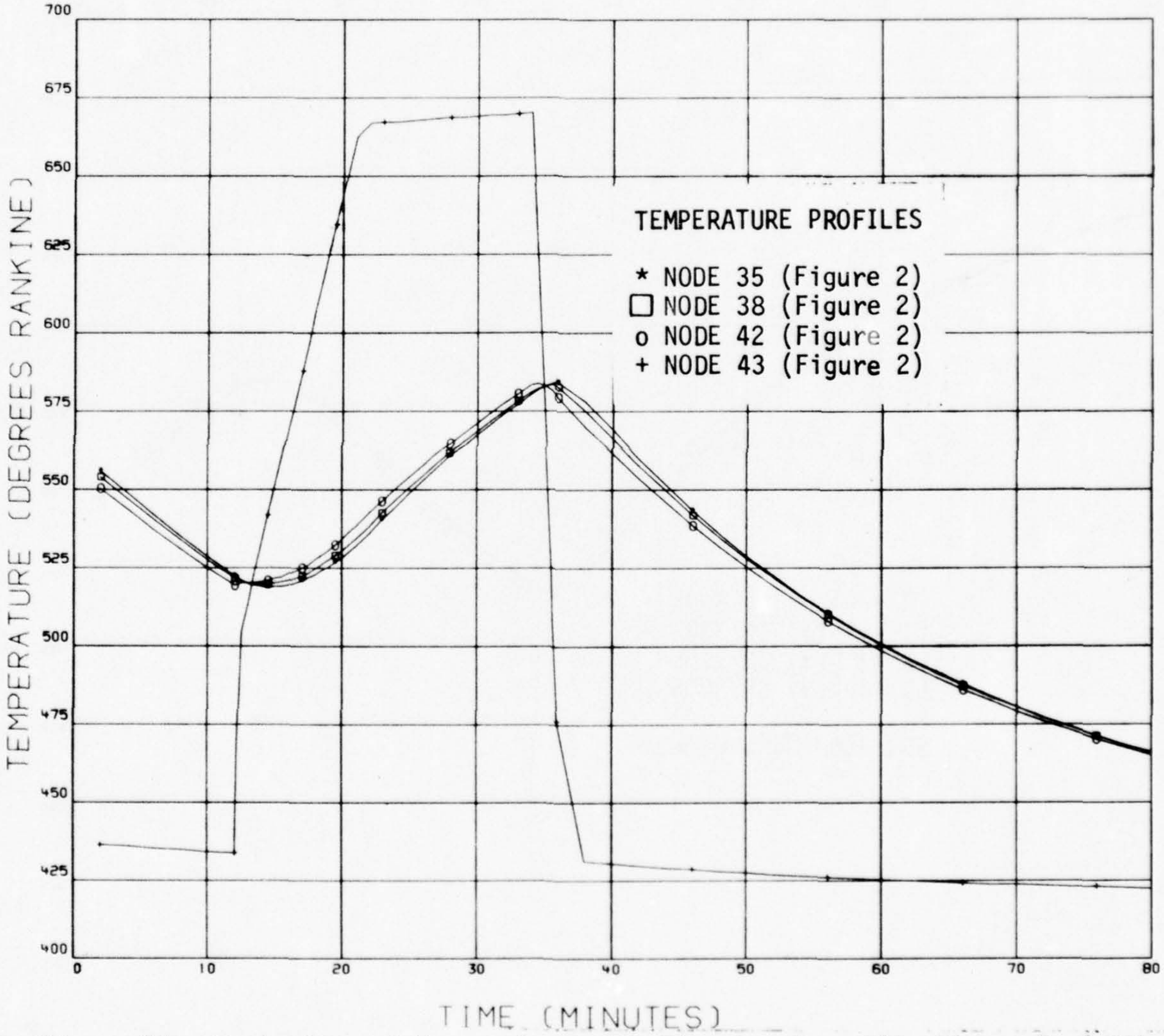


Figure 14. Variations of Temperature with Time for Case III, Figure 6, (with Insulation) at a Cross Section Seven Inches from the Nose

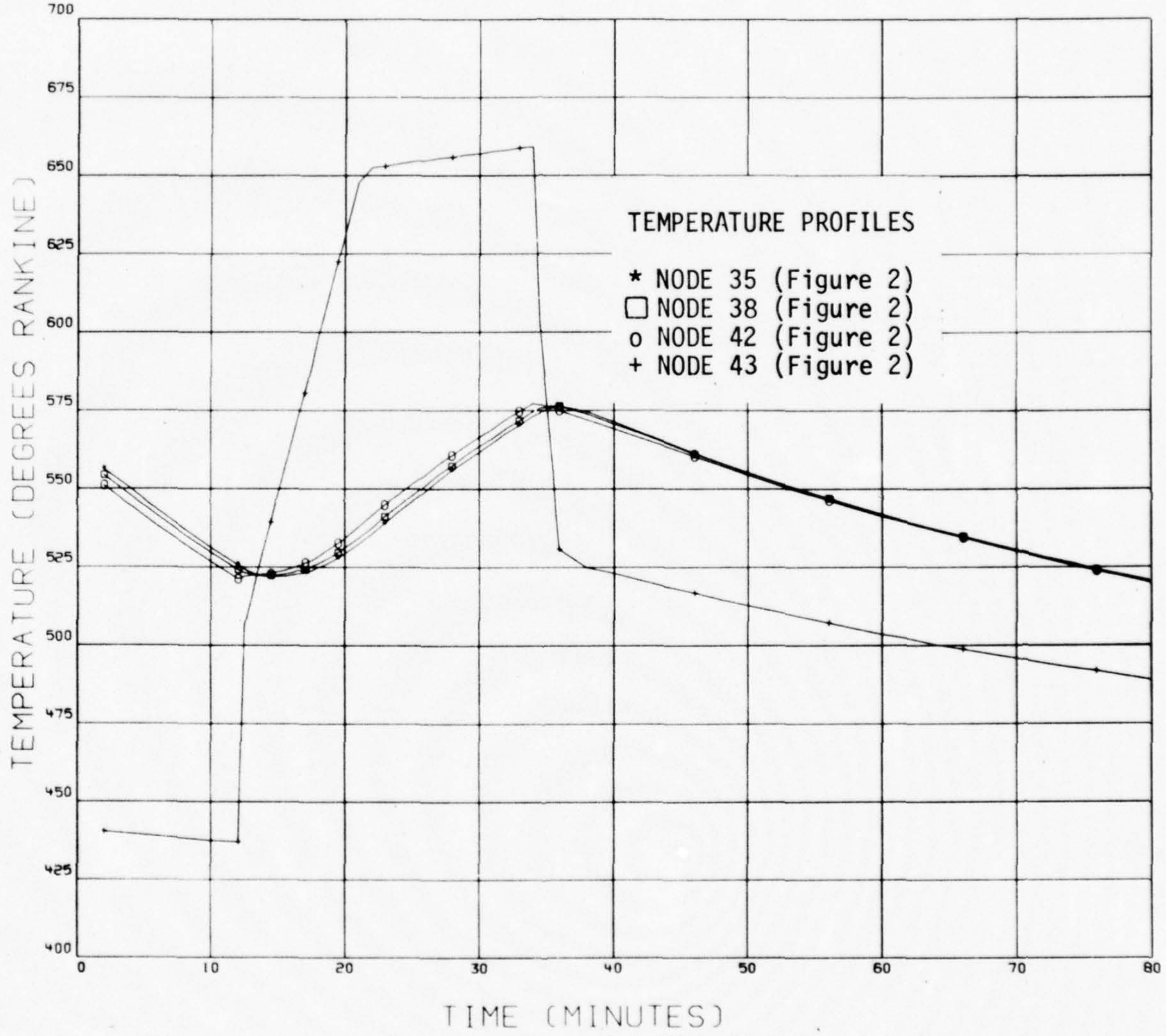


Figure 15. Variations of Temperature with Time for Case IV, Figure 7, [h(Pos., Time)] at a Cross Section Seven Inches from the Nose

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

ANALYSIS OF MISSILE CONTROL SYSTEMS

Prepared by:	Michael E. Warren
Academic Rank:	Assistant Professor
Department and University:	Department of Electrical Engineering University of Florida
Assignment:	
(Laboratory)	Armament
(Division)	Munitions
(Branch)	Mines
USAF Research Colleague:	Major L. D. Berry Haydon Grubbs
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

ABSTRACT

ANALYSIS OF MISSILE CONTROL SYSTEMS
by
MICHAEL E. WARREN

An analysis of a typical missile autopilot is developed and it is shown that the low velocity portion of the trajectory may exhibit instabilities.

Finally, a simulation program for evaluating missile control system design and performance is developed.

INTRODUCTION

The analysis of proposed missile control systems typically involves a full scale digital computer simulation, testing and evaluating the component subsystems both independently and in concert. As the basic requirements for any missile control system are similar, we might expect to find a great deal of commonality and interchangeability between subsystems of different missile control systems. Hence, a vast reduction in the work required of a full simulation could be effected if the simulation were to take advantage of this commonality.

Any missile control system must have at its base a seeker to find the target and give information yielding its relative position or direction from the missile. The seeker angles are then used as an input to the control system which, by changing the position of aerodynamic surfaces, changes the orientation of the missile, aligning it with a preferred direction. Variations on this theme abound, but any control system will have a seeker, actuators, control logic, etc.

A simulation program wherein the different subsystems of a missile control system may be interchanged at will is in development and will be explored in this report. First, a discussion of the most common missile guidance law, proportional navigation is given. Then, a particular missile autopilot is shown and its step response at various portions of a typical flight are shown.

NOMENCLATURE

Symbol	Definition
V	Missile inertial velocity
γ	Flight path angle
ϕ	Seeker line of sight angle
K	Proportional navigation gain
e_γ	Angular error between flight path and los
R	Slant range
R_o	Initial slant range
$e_{\gamma o}$	Initial value of e_γ
γ_o	Initial flight path angle
ϕ_o	Initial line of sight angle
k_1, k_2, k_3	Autopilot gains
k_a, k_5	Servo gains
τ_a, τ_γ	Autopilot time constants
M_δ	Moment due to δ
M_α	Moment due to α
M_{α^3}	Moment due to α^3
I	Inertia of the missile
η	Angle between missile velocity and horizontal
g	Gravity
α_c	Commanded angle of attack
δ	Fin deflection
α'	Measured angle of attack
θ	Missile body angle
λ	Seeker gimbal angle

I. Proportional Navigation Guidance Laws

In proportional navigation, the missile acceleration $V\dot{\gamma}$ normal to the flight path is made proportional to the change of the line of sight ϕ (See Figure 1).

$$\dot{\gamma} = K\dot{\phi} \qquad (1)$$

The measurement of line of sight rate involves differentiating the line-of-sight angles and this introduces noise problems. However, for fixed targets the angular error between flight path and line-of-sight

$$e_{\gamma} = \phi - \gamma \qquad (2)$$

provides the equivalent information. The relation between line-of-sight rate $\dot{\phi}$ and e_{γ} is given by

$$R\dot{\phi} = V\sin e_{\gamma} \qquad (3)$$

and a third relation is

$$\dot{R} = -V\cos e_{\gamma} \qquad (4)$$

Equations (3) and (4) yield the components of velocity along and normal to the line-of-sight, and together with definition of the error (2) and equation (1) represent the kinematics of a proportional navigation guidance law.

The above equations can be integrated to provide an analytical solution of the guidance law. First equation (2) is used to eliminate ϕ, yielding a differential equation in R and e_{γ}. Differentiating (2) and combining with (1) immediately eliminates $\dot{\gamma}$ yielding

$$\dot{e}_{\gamma} = \dot{\phi}(1-K) \qquad (5)$$

which may be combined with (3) to eliminate dependence upon $\dot{\phi}$. Dividing the resultant equation into (4) gives

$$\frac{\dot{R}}{R} = \frac{-1}{1-K}\cot e_{\gamma} \cdot e_{\gamma} \qquad (6)$$

The solution for $K \neq 1$ is given by

$$\log R - \log R_0 = \frac{-1}{1-K}\left[\log(\sin e_{\gamma}) - \log(\sin e_{\gamma_0})\right] \qquad (7)$$

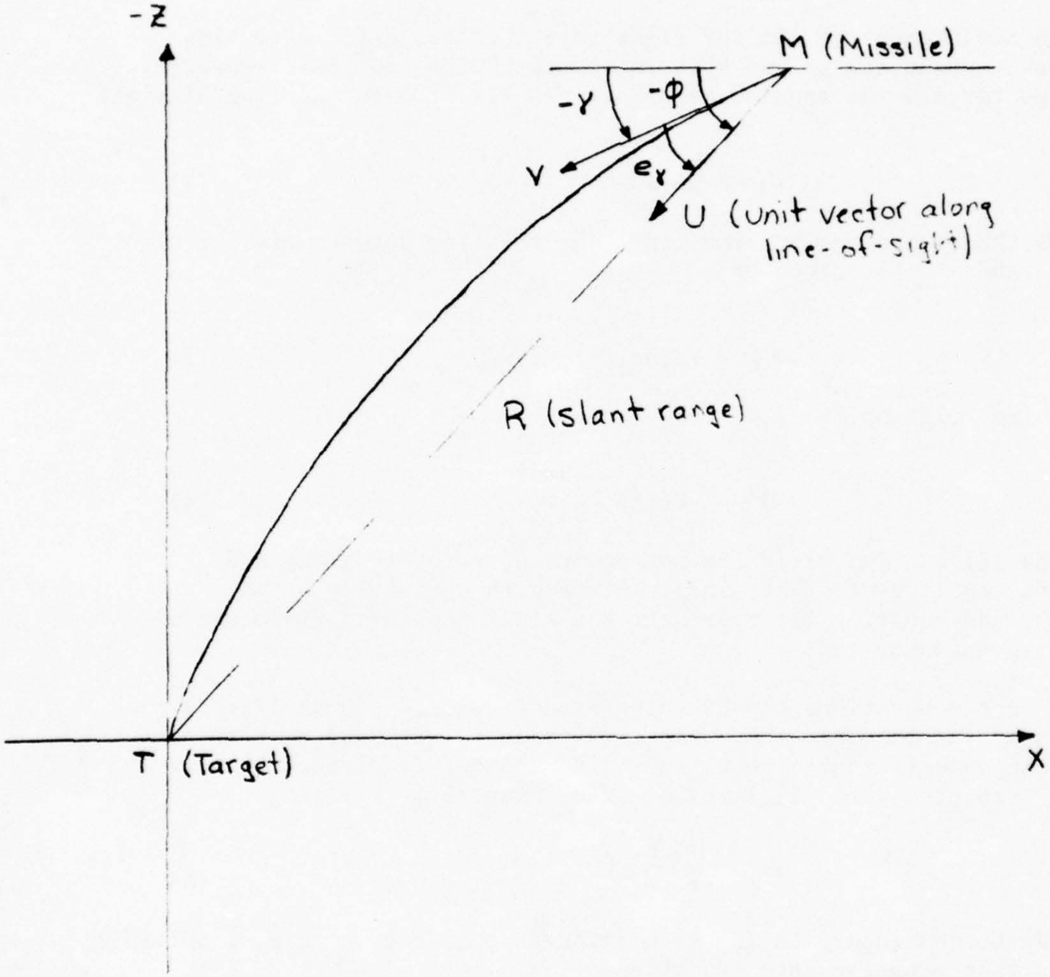


Figure 1. Proportional Navigation Example

By suitable substitution using (1) this may be put in the form

$$\frac{\dot{\gamma}}{\dot{\gamma}_o} = \frac{\dot{\phi}}{\dot{\phi}_o} = \left(\frac{R}{R_o}\right)^{K-2} \tag{8}$$

For $K = 2$, the flight path angle rate is constant, and a constant velocity V will yield a circular trajectory terminating at the target. For $K > 2$ the turning rate goes to zero as the missile approaches the target.

For $K = 1$, equations (1) and (2) yield $\dot{e}_\gamma = 0$ and dividing (4) by (3) yields

$$\frac{\dot{R}}{R} = -\dot{\phi} \cot e_{\gamma o} \tag{9}$$

which may be integrated resulting in

$$\phi = \tan e_{\gamma o} \log\left(\frac{R_o}{R}\right) \tag{10}$$

a logarithimic spiral. If K = 0 then no control is applied and the trajectory is an unguided straight line path.

II. Analysis of a Missile Autopilot

In this section a typical missile autopilot is examined. The seeker is modelled as a first order loop with commanded line of sight rate proportional to seeker error. The overall block diagram for the autopilot is given in Figure 2.

A simulation of the control system was accomplished assuming a horizontal engagement plane (η=0). The pitching moments due to α, α^3 were deemed small and also ignored. Reference to Figure 2 indicates that there are two basic inputs to the control system ϕ and $\dot{\phi}$. For this analysis we have chosen to examine the system relating θ to ϕ, i.e. a pursuit navigation mode.

For this phase of the trajectory the control system is a unity feedback type with forward loop transfer function.

$$G = \left(\frac{K_2 V + K_3 \tau_\gamma S}{V s^2}\right) \frac{\frac{K_s M_\delta}{I}\left(s + 1/\tau_\gamma\right)}{s^3 \frac{\tau_s}{\tau_\gamma} + s^2(1+\tau_s) + s\left(\frac{K_s M_\delta}{I}\tau_a + \frac{1}{\tau_\gamma}\right) + \frac{K_s M_\delta}{I}} \tag{11}$$

Parameter values are found to be

$$\tau_s = 0.005$$
$$\tau_a = 0.03$$
$$\tau_\gamma = 1$$

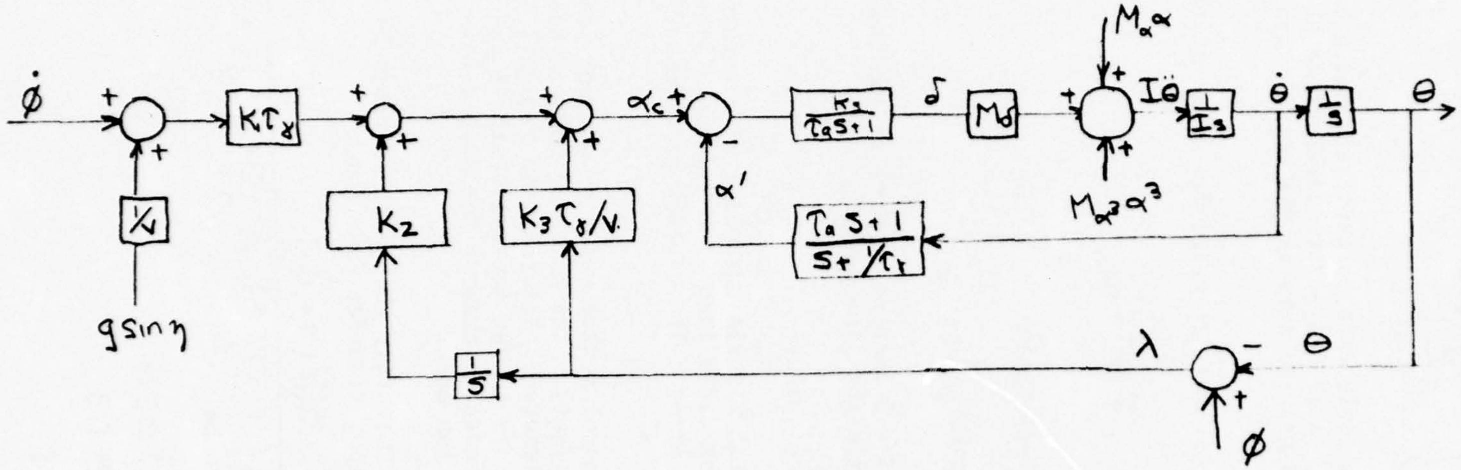


Figure 2. Missile Control System

$$\frac{K_s M_\delta}{I} = 2182$$

resulting in a forward loop transfer function

$$G = \frac{10V + 304.8s}{Vs^2} \cdot \frac{2182(s+1)}{0.005s^3 + 1.005s^2 + 6646s + 2182}$$

The step response of the closed loop system was determined for values of missile velocity ranging from 100 m/sec to 600 m/sec in increments of 100 m/sec. These responses are shown on the following charts, Figures 3 thru 8.

From the graphs of the step responses, it is clear that the high velocity behavior is dominated by a single real pole. As V decreases however, oscillatory effects of a complex pair of poles becomes increasingly evident.

The closed loop system poles have been found for various values of V. These are presented in Table 1 following the graphs of the step response. As V decreases the complex pair of open loop poles at -34.75 ± 33.85j move toward the right half plane, crossing over at approximately V = 68 m/sec. For velocities less than this the control system will be unstable. A root contour showing the variations of these roots with V is also plotted in Figure 9.

III. Computer Simulation

A computer program to allow a sophisticated simulation of missile control system performance is in the developmental stage. To gain the required accuracy for performance determination, it was felt a six degree of freedom simulation was needed. Rather than start from scratch and develop the data handling routines as well as routines simulating missile behavior, an in-house simulation package was used as a basis. The entire data handling structure of the simulation was then salvaged. The computer program in development is based upon a package put together for the Air Force several years ago (see reference 1). Data storage, retrival and manipulation is handled under the control of executive routines. Particular missile control system subsystems are modeled in specific subroutines provided by the user. Thus the user has at his command the ability to construct the simulation just as he might design an actual control system, using functionally oriented building blocks.

At this time subprograms exist to compute gravitational and coriolis forces, model steady state winds as well as provide randum gusts, compute atmospheric parameters, and determine the transformations required between different coordinate systems. Further models of a phase comparison terminal seeker, an inertial guidance platform, an acceleration and rate autopilot, and a proportional navigation steering computer are available for use. Additional programs to compute aerodynamic forces and moments,

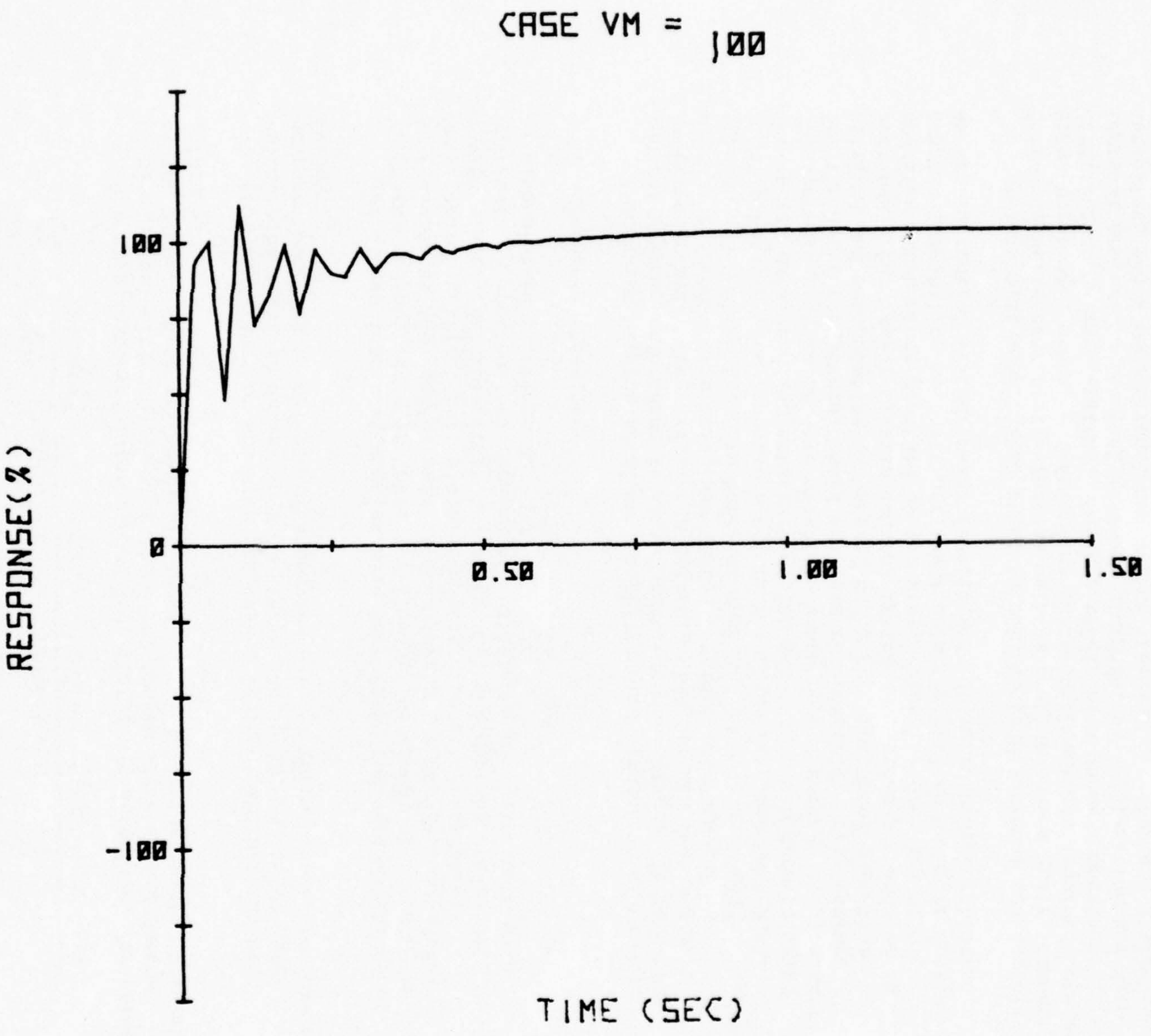


Figure 3. Step response of missile autopilot (V = 100 m/sec)

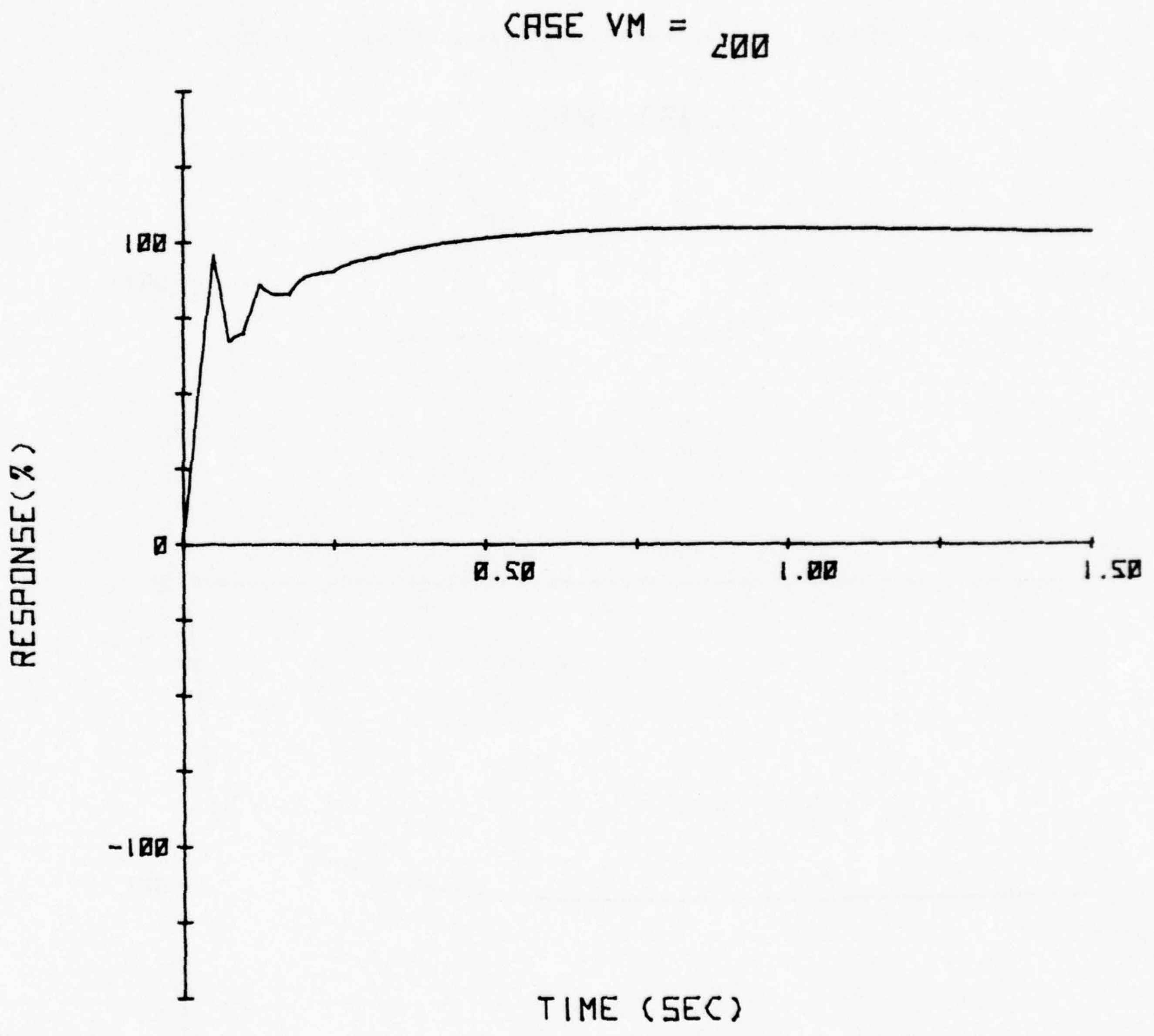


Figure 4. Step response of missile autopilot (V = 200 m/sec)

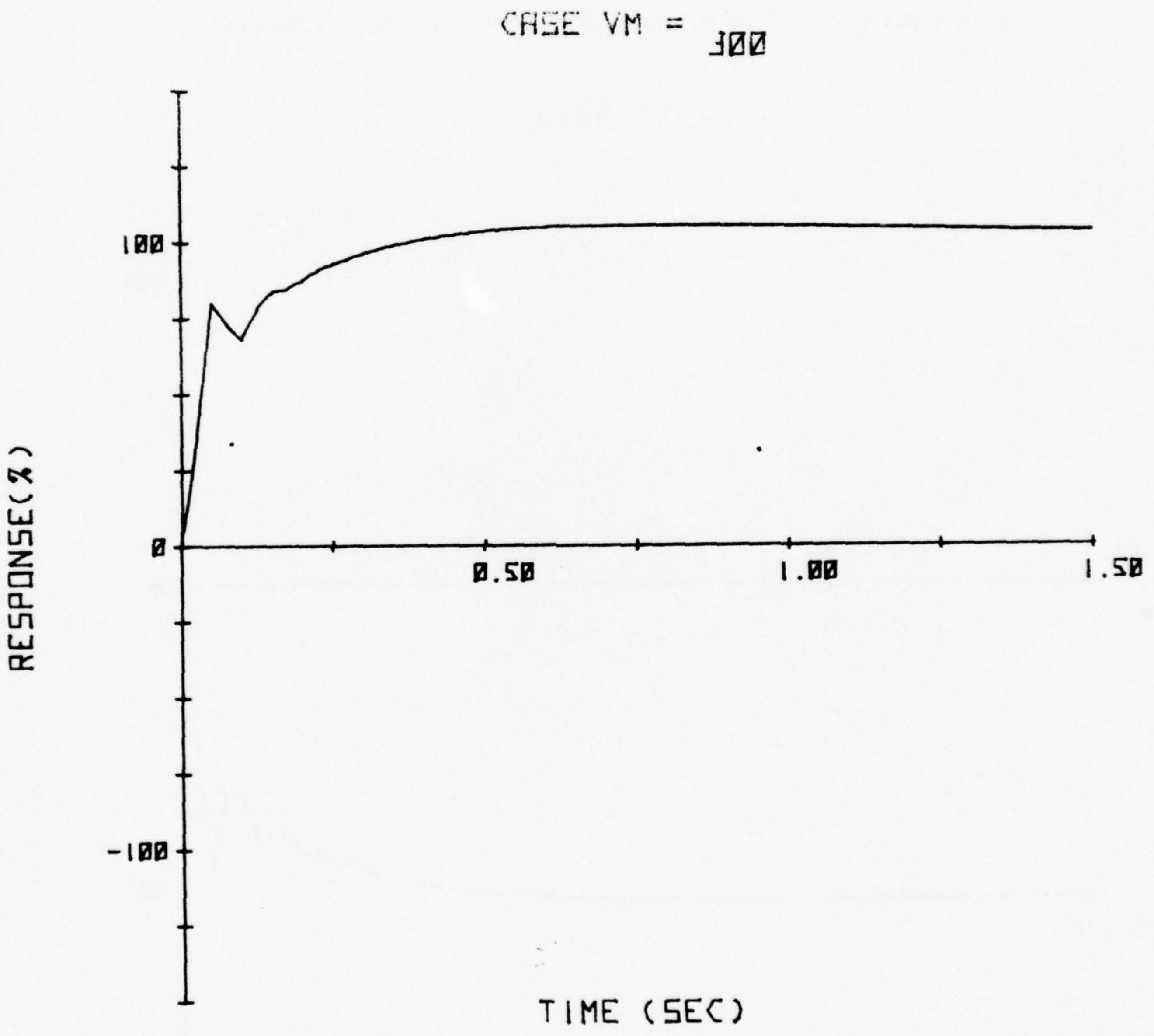


Figure 5. Step response of missile autopilot (V = 300 m/sec)

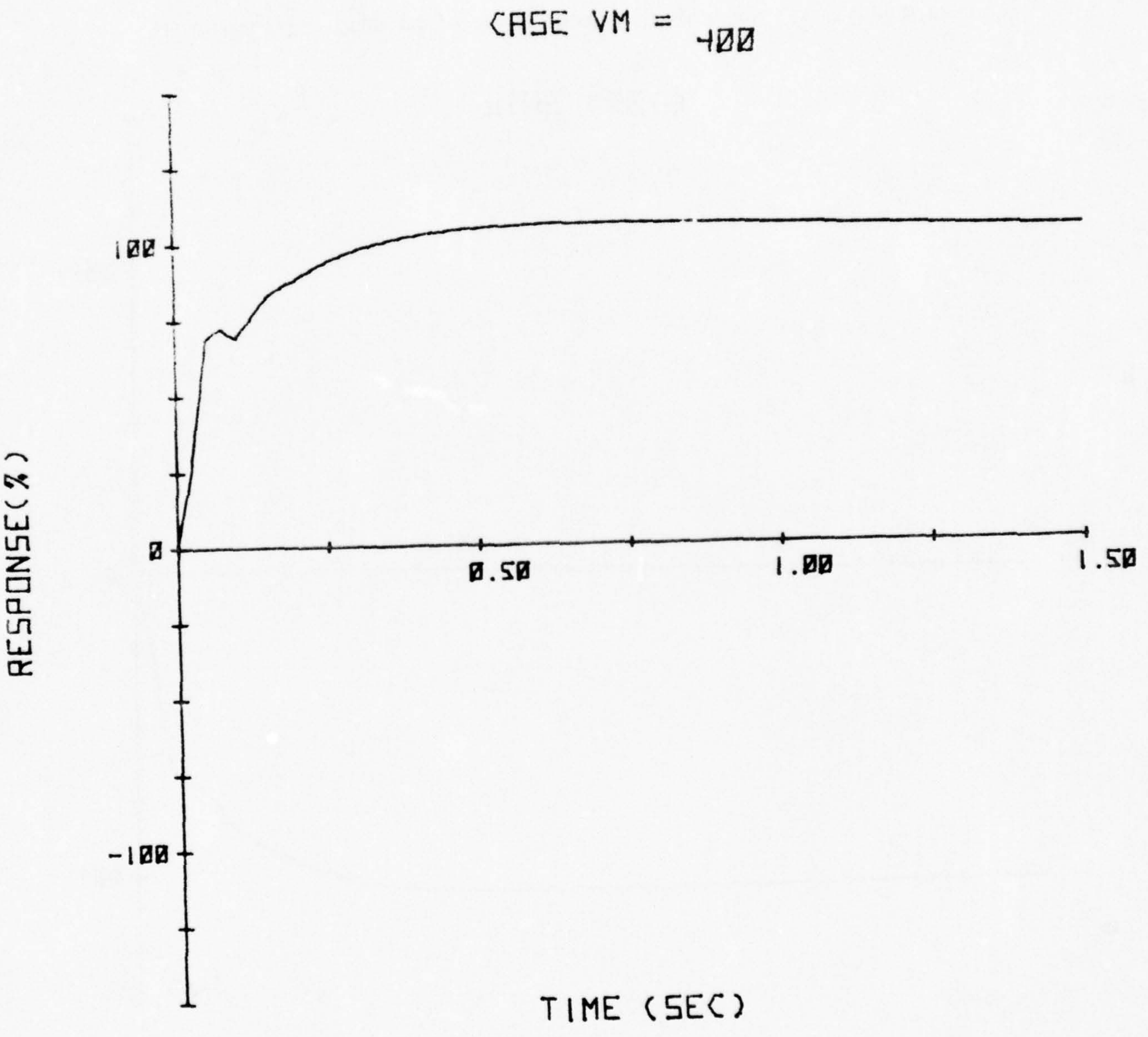


Figure 6. Step response of missile autopilot (V = 400 m/sec)

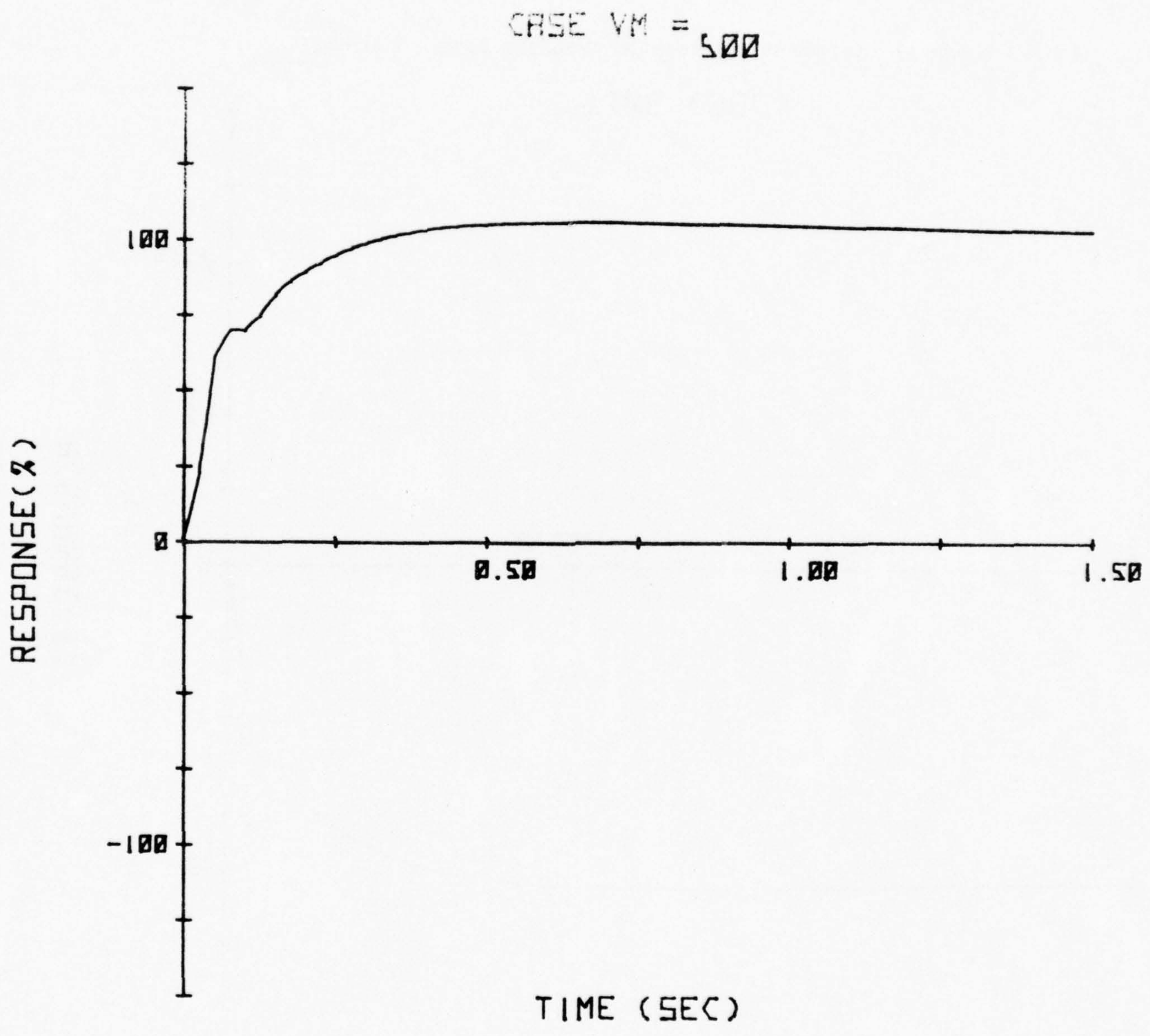


Figure 7. Step response of missile autopilot (V = 500 m/sec)

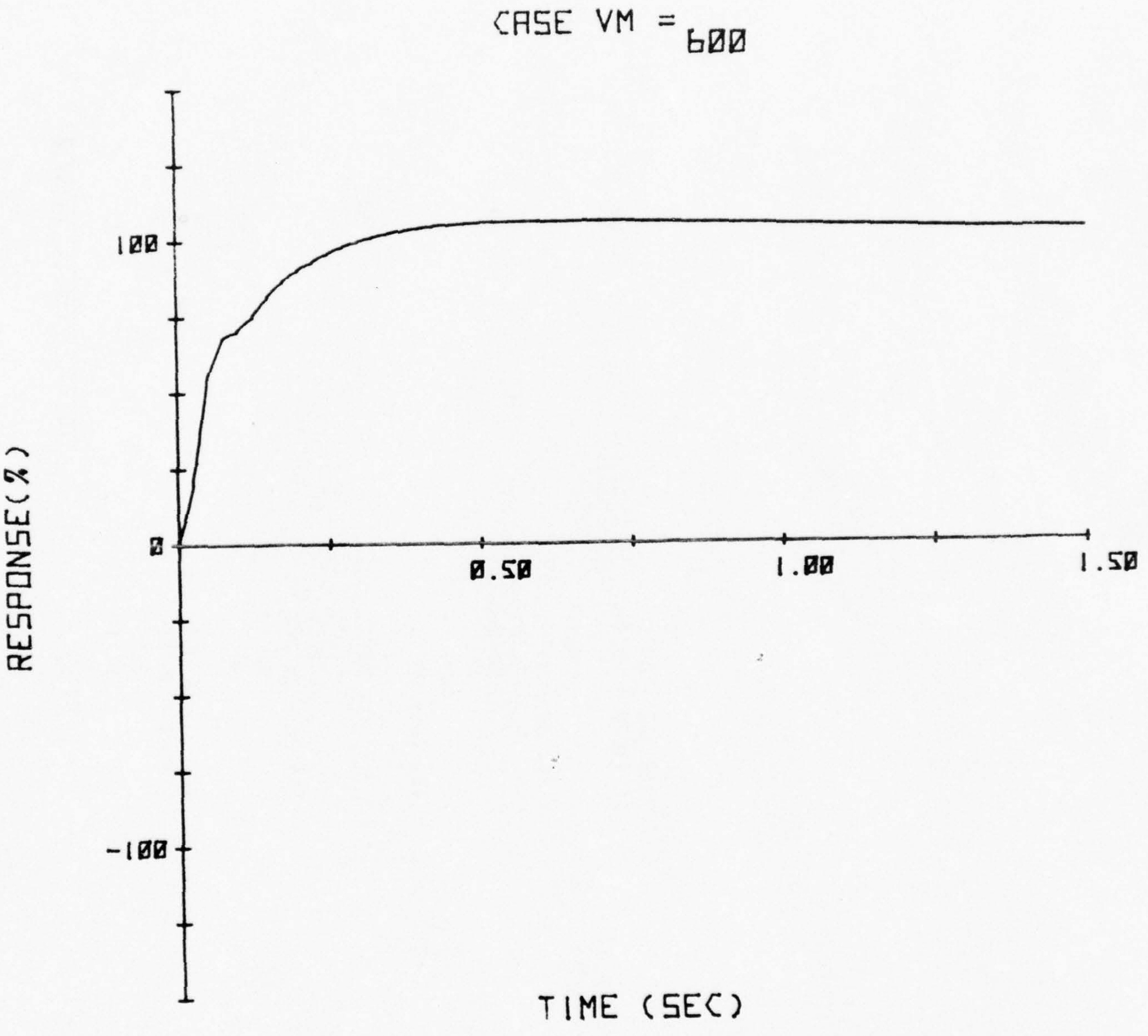


Figure 8. Step response of missile autopilot (V = 600 m/sec)

Table 1

Closed loop system poles

Missile Velocity (m/sec)	Poles
500	-1.13
	-6.28
	-26.66 ± 50.81j
	-130.61
250	-1.15
	-4.22
	-20.72 ± 73.62j
	-154.20
166	-1.17
	-3.17
	-16.00 ± 83.12j
	-164.66
100	-1.24
	-2.05
	-8.56 ± 97.34j
	-180.60
50	-1.14 ± 0.34j
	4.60 ± 121.37
	-207.92

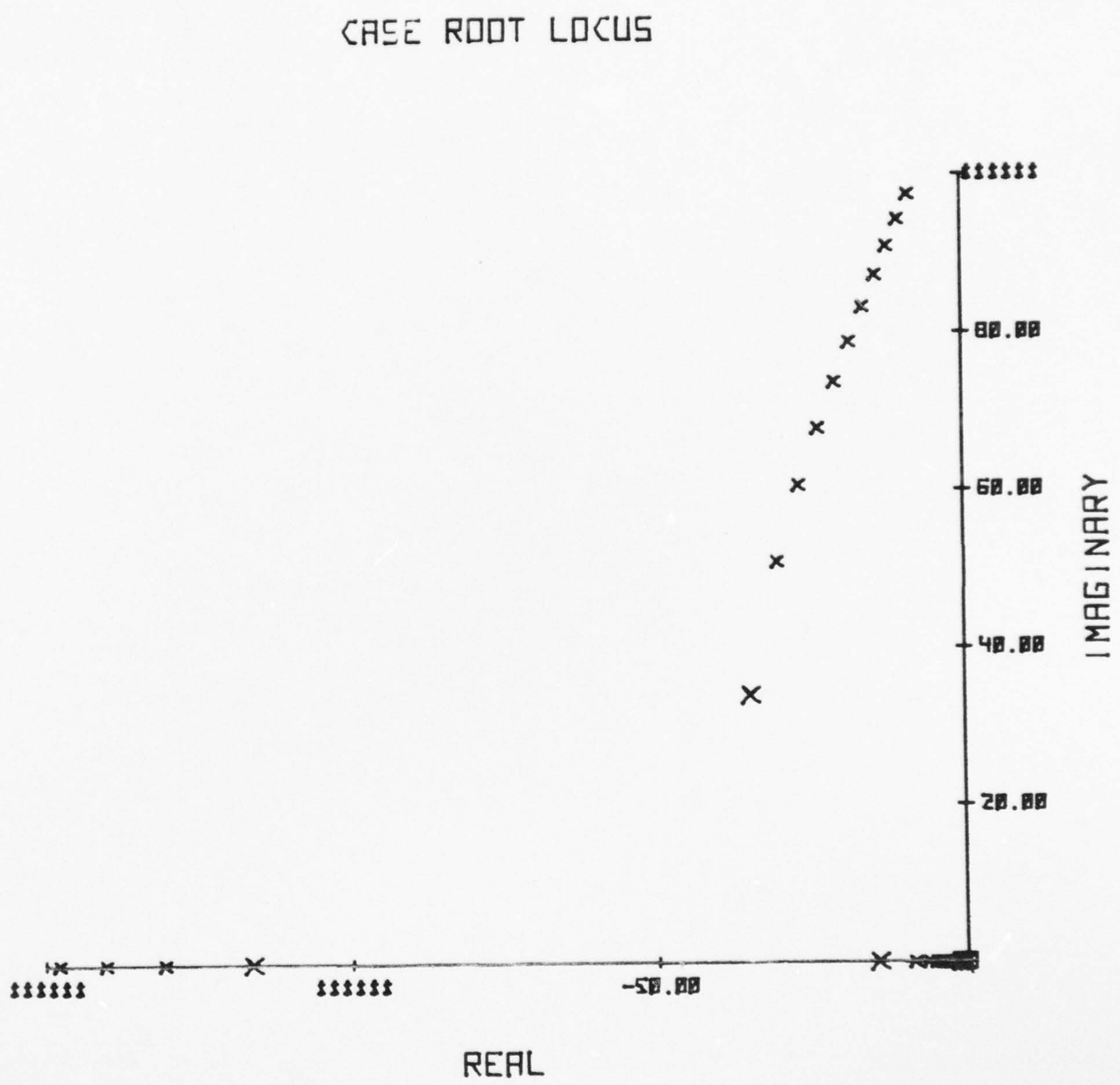


Figure 9. Migration of Autopilot closed loop poles with changing velocity

determine aerodynamic coefficients, translational and rotational dynamics exist. The resulting equations of motion are integrated by a sophisticated combination Adams-Moulton/Runge-Kutta scheme. The particular subprograms mentioned above are all missile dependent would have to be modified for use in any particular simulation.

The program as it now stands does not allow for the plotting of data. Such graphical displays will be of great value and will be incorporated into the simulation package.

References

1. Diesel, J.W., et al. Modularized Six-Degree-of-Freedom Computer Program, Volume 1, Wright-Patterson Air Force Base, Ohio

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

DIGITAL AUTOPILOT DESIGN

Prepared by:	J. N. Youngblood, PhD.
Academic Rank:	Associate Professor
Department and University:	Department of Aerospace and Mechanical Engineering University of Alabama
Assignment:	
(Laboratory)	Arnament Laboratory
(Division)	Digited Guided Weapons
(Branch)	Systems Analysis and Simulation
USAF Research Colleague:	Major K. A. (Al) Gale
Date:	August 15, 1975
Contract No:	F44620-75-C-0031

PREFACE

This report comes as a result of a ten week effort that the author and Dr. Jim Delansky undertook at the request of the Systems Analysis and Simulation Branch of the Guided Weapons Division of the Armament Laboratory, AFATL, USAF. Under the ASEE-USAF Faculty Fellowship Program 1975, we were asked to prepare a concise, qualitative survey of those areas of digital controller design that we felt to be most important to the branch for their particular effort. Those areas on which I chose to concentrate are as follows:

(1) error sources in digital autopilots

(2) design of digital autopilots using discrete optimal estimation and control.

Clearly, both of the topics are fully covered in the literature, which is liberally referenced herein. However, this paper discusses the topics in the light of applications to guided weapons.

In most cases the ideas expressed in this document are the result of condensing and compacting the work of others to form a concise package. Complete developments of the individual subjects exist in the literature which is referenced.

I. INTRODUCTION

1.1 Background

For a number of years the design philosophy for guided weapons has been roughly (and perhaps somewhat out of order) the following:

(1) Design the airframe for a warhead with particular attention to structural and aerodynamic properties.

(2) Select a target seeking system depending on the type target and delivery system but rather independently of the airframe.

(3) Design a guidance or trajectory control concept in keeping with the target information to be received and based upon expected target dynamics.

(4) Design stability control loops for the unguided airframe.

(5) Implement three and four jointly or separately with a hard-wired analog autopilot, having at most several different configurations for different flight/guidance regimes.

A different concept has recently evolved for classes of guided weapons. The newer concept is roughly as follows:

(1) Obtain (not necessarily from new design) a class of airframes to be guided.

(2) Select a family of complementary target seekers.

(3) Design a set of guidance concepts, each of which may, or may not, be tied to a given target sensor.

(4) Design a digital autopilot with a specified architecture that may be programmed for any of the classes of airframes, directed by

any of a subclass of seekers, for any of a subclass of guidance laws to provide a stable guided vehicle.

Such a concept is highly dependent on the final design step. This paper discusses some of the aspects of the design and programming of a digital autopilot to function for a given choice of items 1, 2, and 3.

1.2 Requirements of the Autopilot

The autopilot, digital or otherwise, performs two distinct functions. The stability augmentation or "inner loop" function is that of actively controlling the airframe to present for guidance input a system with the proper selected aerodynamic characteristics. The inner loop requirement is primarily influenced by the aerodynamic properties of the airframe and the flight regime.

The second function of the autopilot is to generate guidance commands for the stabilized airframe based on target seeker information. This "outer loop" control function is determined by target dynamics, seeker dynamics, and stabilized weapon dynamics. While the two functions are required of every autopilot, there are design approaches, such as the error minimization type, which tend to merge them.

There are implied requirements of the autopilots, not explicitly mentioned in the inner loop, outer loop discussion. The autopilot must be able to extract meaningful information from the attitude sensors and from the guidance seeker. This process entails a certain amount of filtering or state estimation. The autopilot must sense switching times for gain switching between various flight regimes and accomplish this function. In sophisticated systems where parameter estimation and adaptive control are employed, these are also performed in the autopilot. Other operations could be considered requirements of the autopilot, if the autopilot had the capacity and flexibility to perform them.

1.3 Fundamentals of the Digital Autopilot

1.3.1 Objective

The objective of this study is to produce a concise document covering selected areas of digital controller design applied to a class of guided weapons and to be used by personnel already trained in analysis and simulation techniques of continuous autopilot design. This particular paper is one part of a two part set to accomplish the objective. While each part is self-contained, the basic theory of discrete system design is surveyed in the other part of this set. This part is directed toward the following:

(1) error sources in digital autopilots.

(2) optimal control and estimation using digital autopilots.

1.3.2 Fundamental considerations in digital autopilot design.

In considering the use of digital autopilots in guided weapons, there are several features that make this approach attractive. There is no doubt, for the computational load required to replace the analog autopilot, that such a replacement can be made digitally. Moreover the greater programming flexibility of the digital instrument is an important advantage for the modular concept. Consequently, it is appealing to design a discrete autopilot that is general enough in structure for every airframe/guidance combination and to program it to mimic a satisfactory analog autopilot in each case.

There are several reasons why a fresh approach, other than copying an analog version, can pay dividends. For one, the ease with which complicated logic can be performed opens new options for control. For another, the capacity to store tabular data for recall offers a wealth of possibilities in the areas of parameter and gain storage, parameter estimation, and adaptability of control. Finally, the techniques of state estimation and optimal control are extremely well suited for implementation by a digital processor. There are, of course, prices to pay for additional computation, but the possibilities exist. Moreover, there is no reason why stability augmentation, or other routine control, cannot be performed by analog, leaving the digital instrument more capacity for guidance.

1.3.3 Digital design approaches

The design of a discrete time controller for a continuous time plant is usually accomplished by either discretizing the plant model and using digital design principles or by using continuous time design principles and discretizing the resulting controller. In the former case when discrete design is done in a conventional manner, the general area of sample-data control system design applies. The design procedures are not discussed in this report; however, they are an important part of the subject matter. Certain aspects of the other two cases, optimal discrete design and discretization of analog models, are covered here.

2. ERROR SOURCES IN DIGITAL AUTOPILOTS

2.1 Formula Error in Discretization

2.1.1 Continuous system modeling

If the design concept of the digital autopilot is that of transforming a continuous autopilot, there are errors introduced by the approximation of the differential equations by their discrete counterparts. Such errors depend on the formula that is used for discretization. Perhaps the most crucial decision in designing the autopilot by this approach is the selection of the transformation. The choice yielding the least formula error depends almost exclusively on how the error is defined.

The customary error criteria may be grouped in the following warp:

(1) There should be minimum (in some sense) difference between the time responses of the continuous system and the discrete system for a specified input, or class of inputs.

(2) There should be minimum (in some sense) difference between the frequency responses of the two systems.

The error in the time response sense is input dependent. For this reason a transformation for a system yielding a suitable error for impulse or step response could yield an intolerable error for inputs encountered in actual operation. Since the true inputs to the autopilots can be described more readily in terms of their frequency content, deterministically or statistically, there is an impetus to use an error criterion based on frequency response matching. There is a considerable amount of published work in the area of modeling continuous filters by digital filter to preserve frequency characteristics. (2-7) There is no essential difference in the application here.

2.1.2 The continuous to discrete transformation by numerical integration

There are three general approaches to the problem of discrete-modeling of continuous systems. Firstly, the derivatives may be approximated by differences, as is common in computer simulation of differential equations. Secondly, the system may be chosen to have invariance of the time series output for a specified input. Thirdly, the frequency response of the system may be preserved. While it is true that the third approach is identical to the second for a sinusoidal class of inputs, there is enough difference in philosophy to warrant the distinction.

If the differential equation is discretized by the first backward difference

$$\frac{dx}{dt} \qquad \frac{x(n)-x(n-1)}{T}$$

the transformation from s to z is

$$s = \frac{1-z^{-1}}{T}$$

where T is the sampling period. The first forward difference

$$\frac{dx}{dt} \qquad \frac{x(n+1)-x(n)}{T}$$

yields the transformation

$$s = \frac{z-1}{T}$$

The transformation of backward difference preserves the stability of the system by mapping poles of H(s) with negative real parts into Z plane poles with magnitude less than one. The forward difference transformation maps a pole p in the s plane into the z plane pole pT + 1, which lies in the stability region, only if the sampling period is smaller than $\frac{-2 \text{ Re } \{p\}}{|p|^2}$. Hence all poles with negative real parts can be mapped inside the unit circle if the sampling frequency is sufficiently high.

Digital systems designed in this manner for on-line operation usually require a sampling frequency that is relatively high to represent the continuous system for any, but very low, frequency inputs. The tradeoff between sampling frequency and the amount of computation to be performed makes this approach unfavorable.

2.1.3 The continuous to discrete transformation by input invariance

The design of an automatic controller to achieve a satisfactory response to a standard input is a customary approach. Consequently, it is logical to transform a continuous controller to a digital controller that has the same response, point-by-point, as the continuous controller, to a given standard input. Such an approach is called the principle of irput invariance or point-by-point design. If a linear, constant coefficient system has an impulse response g(t) and an input x(t) the output u is given by $u(t) = \int_{-\infty}^{\infty} g(\tau) \times (t-\tau)d\tau.$ (2.1)

If the input x(t) is sampled and fed to a linear, constant coefficient discrete system with pulse response h(t), the output of this system is

$$v(nT) = \sum_{k=-\infty}^{\infty} h(kT)\ x\ (nT-kT) \qquad (2.2)$$

If the discrete system h is to have the same output as the continuous system g, point by point, then

$$\sum_{k=-\infty}^{\infty} h(kT)\ x\ (nT-kT) = \int_{-\infty}^{\infty} g(T)\ x\ (nT-\tau)d\tau. \qquad (2.3)$$

for all n.

The above expression is, of course, input dependent. For an impulse input, the above reduces to

$$h(nT) = g(nT) \qquad \text{for all } n.$$

For unit step invariance and causal systems,

$$\sum_{k=0}^{\infty} h(kT) = \int_{0}^{t} g(\tau)d\tau. \qquad (2.4)$$

Since the input invariance approach is a function of the input, such a method leads to errors (point-by-point) when the input is not the type for which the system is designed. Where the continuous system was chosen on the basis of a transient response (for example, overshoot and damping for a step input) to a standard input, such a procedure might be appropriate.

Consider the impulse response of a discrete system h*(t) that was selected to be impulse invariant with respect to a continuous system g(t). [1]

$$h^*(t) = \frac{1}{T} \sum_{n=-\infty}^{\infty} g(t)\ e^{jnw_s t} \qquad (2.5)$$

It is easily shown that the pulse transfer function of the discrete system is related to the transfer function of the continuous system as

$$H(z) = H^*(s) = \frac{1}{T} \sum_{n=-\infty}^{\infty} G(s+j2\Pi n/T). \qquad (2.6)$$

The modulation effect due to sampling, called aliasing, causes the frequency response of the discrete system to be comprised of a fundamental frequency response component similar in form to G(jw) and harmonic components, also similar to G(jw), but centered at multiples of the sampling frequency. This version of the discrete controller is not satisfactory for large bandwidth inputs or inputs with noise, even if low pass filtering is used for continuous signal reconstruction, because of folding back of the high frequency input components due to aliasing.

2.1.4 The continuous to discrete transformation by modelling frequency response.

The most widely examined means of continuous to discrete system transformation is that used in digital filter design. In these applications frequency response specification is the objective. Desired is a one-to-one transformation of s into z that maps the s plane region of stability into the z plane region of stability and preserves as much as possible the frequency characteristics of the continuous systems

The relationship between s and z

$z = e^{st}$ or $s = T^{-1} \ln z$

may be expanded as follows into log series

a. $s = T^{-1} \sum_{n=1}^{\infty} (1-z)^{n}/n \qquad |z-1| <1 \qquad (2.7)$

b. $s = T^{-1} \sum_{n=1}^{\infty} (1-z^{-1})^{n}/n \qquad |z| >1/2 \qquad (2.8)$

c. $s = T^{-1} \sum_{n=0}^{\infty} 2\left(\frac{z-1}{z+1}\right)^{zn+1} 2n+1)^{-1} \qquad |z| >0 \qquad (2.9)$

If s is approximated by the first term of series (a) the first forward difference transformation

$$s = \frac{1-z}{T} \qquad (2.10)$$

is obtained. The first term approximation of series (b) yields the first backward approximation

$$s = \frac{1-z^{-1}}{T} \qquad (2.11)$$

The first term of series (c) yields

$$s = \frac{2}{T} \quad \frac{z-1}{z+1} \qquad (2.12)$$

A transformation known as bilinear, which is single valued, preserves the stability properties of the system and in many cases leads to suitable frequency properties of the digital system. The frequency response errors due to this transformation are extremely well documented in the literature pertaining to digital filtering and not repeated here. (2-7)

2.2 Effects of Finite Word Length and Quantization

2.2.1 Error sources

The digital autopilot operates on continuous signals from a variety of sensors using computational algorithms to produce sequences of discrete commands for the control actuator. Consequently, a limitation is inherently imposed on the accuracy of the autopilot by the granulation of the analog-to-digital conversion and by the retention of only a limited number of bits in the algorithm. The errors attributed to finite word length sources can be grouped as follows:

(1) quantization of the analog measurement signals,

(2) finite register length for storing the coefficients of the autopilot algorithm,

(3) accumulation of truncation errors in the processed signals.

The magnitude of the error produced by these effects is influenced by the mechanics of autopilot processing, for example:

(1) the type of transfer function realization adopted for the implementation of the autopilot,

(2) the type of arithmetic used in the algorithm, fixed point, floating point, integer, fraction, etc.,

(3) the type of approximation used to represent numbers with a finite number of bits, chopping or rounding,

(4) the manner that negative numbers are represented.

A substantial amount of study on the topic of errors due to finite word length has been published, primarily in the literature devoted to digital filtering. (8-13) The general object of this type of study is to define an output error due to a combination of finite word lengths inaccuracies and to propagate the statistics of the error through the algorithm. Many of these papers are referenced and borrowed from. The object here, however, is not one of quantifying the error, but one of discussing the mechanisms that produce it.

2.2.2 Input Sequence Quantization

The input signal converter quantizes the analog signals from the sensors at levels separated by a level dependent step size. In the simplest and most common case, the signal is sampled at a uniform rate and quantized with equal step size. Some sensor models have digitized outputs and require no additional conversion other than scaling. Systems operating with these sensors may be thought of as analog to digital systems with the conversion applied at the sensor. If, as is the usual case, the word length of the A/D converter is shorter than the register length of the autopilot, the ultimate precision of the controller may be determined by the initial conversion.

The usual approach for the determination of the input quantization effects is to define an error due to initial quanitization and to propogate the error statistics through the filter (autopilot). The spectrum of the error at the output due to input quantization is related to the spectrum of the same error at the input by familiar

$$\Phi_{\infty}(w) = H(z)H(z^{-1})\Phi_{ii}(w) \qquad (2.13)$$

where H(z) is the pulse transfer function of a linear, time invariant autopilot.

2.2.3 Accumulation of error due to roundoff

The mechanism by which error is introduced in finite bit retention is a function of the type of arithmetic, the nature of the algorithm, the type of truncation, and the number representation. For this reason a discussion of the associated error is rather involved. The object here is to discuss each of the above items, its effect on the introduction of error at each step, and the interrelation of one with another. In doing so we generally follow Appenheim and Weinstein (3).

Digital hardware under consideration uses a binary representation with either fixed or floating point arithmetic. For fixed point arithmetic the binary point in each register is fixed. If the registers in a processor have corresponding binary points, addition is accomplished independently of the location of the binary point. The process of multiplication in fixed point, however, requires a knowledge of the binary point location. For floating point arithmetic a number is represented by a characteristic and Montissa in the form

$$N = 2^{c} \cdot M$$

Disregarding negative numbers momentarily, addition of floating point numbers accomplished by adjustment of the characteristic and addition of the Montissas; multiplication is accomplished by addition of the characteristics and multiplication of the Montissas.

For fixed point arithmetic processing, the data is usually scaled, so that all numbers remain less than unity magnitude. If overflow is prevented, addition does not require any truncation of digits, and the product of multiplication may be easily truncated by the elimination of the least significant bits.

For floating point arithmetic the montissa of the number representation is less than unity. Addition of two numbers is accomplished by shifting digits in the smaller number until its characteristic corresponds to the larger, then adding the montissas. In shifting the montissa for addition the least significant bits may be eliminated. In multiplication with floating point numbers elimination of the least significant bits in the product montissa is necessary, as it was with fixed point numbers.

Fixed point arithmetic with fractions requires termination of digits in multiplication, but not addition. Floating point arithmetic requires termination of digits in both multiplication and addition, but has a greatly expanded range of operation.

The choice of method by which negative numbers are represented influences the accumulated roundoff error of the processor. The reason is that roundoff and chopping increases or decreases the magnitude of a negative number depending upon how the number is represented with digits.

It has become customary to represent negative fixed point numbers and the montissas of negative floating point numbers with a one's-complement or two's-complement representation. If a number N is represented by b+1 bits (one to the left of the binary point and b to the right), then negative N is represented in two's complement notation as the binary number equivalent to 2-N. The one's-complement representation of negative N is the binary equivalent of $2-2^{-b}-N$, the number $2-2^{-b}$, being the largest possible number that can be held in the register. The one's- and two's-complement representation of negative numbers is considerably more intuitive in terms of the conjugation of bits. The one's complement representation of -N is achieved by conjugation of all bits in N.

N = 0.10011010 -N = 1.0100101

The two-s'complement representation of -N is achieved by conjugation of all bits in N except the least significant "one" and all less significant "zeros".

N = 0.1011010 -N = 1.0100110

The representation of negative numbers in one's- or two's-complementary notation is done for the purpose of increasing the effectiveness of the processing and not for roundoff effect. Yet the roundoff error is a function of the representation.

As indicated, the effect of imposing a finite word length depends upon whether fixed or floating point arithmetic is used and the manner in which negative numbers are represented. The approximation error is also a function of whether the truncation is carried out by chopping or by rounding. In chopping the bits following the least significant bit are disregarded. In rounding, a 1 or a 0 is added to the chopped number, depending upon whether the first chopped bit was a 1 or a 0.

For positive fixed point numbers truncated by chopping, the error due to truncation

$$e = NT-N$$

is obviously non-positive. If a b_1 bit number is chopped to b_2

bits the truncation error is bounded as follows:

$$-(z^{-b2} - z^{-b1}) \leq e \leq 0 \qquad (2.14)$$

Chopping of a one's-complement negative fixed point number results in a non-negative magnitude error, or a non-positive truncation error for the number bounded as follows:

$$0 \geq e \geq -(z^{-b2} - z^{-b1}) \qquad (2.15)$$

On the other hand chopping of a two's-complement negative fixed point number results in a similar non-negative truncation error

$$0 \leq e \leq (z-^{b2} - z^{-b1}) \qquad (2.16)$$

Thus the process of chopping for truncation is biased, depending on the sign and representation of the number.

Rounding of fixed point numbers results in errors whose statistics are independent of the sign and representation of the number. The bounds for roundoff error are

$$-1/2(2^{-b2} - z^{-b1}) \leq e \leq 1/2(2^{-b2} - z^{-b1}) \qquad (2.17)$$

It is clear that rounding errors are unbiased for both number representations. This is a property which can be extremely important when accumulation of error is to be avoided.

Truncation errors in floating point numbers occur in the montissa and such montissa errors are exactly the same as for fixed point numbers. The total error in the number is, however, the product of montissa error and 2^c. The essential information of this section is summarized in Tables 1 and 2.

	ADDITION	MULTIPLICATION
FIXED POINT	NO ERROR	TRUNCATION ERROR
FLOATING POINT	TRUNCATION ERROR	TRUNCATION ERROR

TABLE 2.1 ERRORS IN ARITHMETIC OPERATIONS

	POSITIVE NUMBER	ONE's COMPLEMENT NEGATIVE	TWO'S COMPLEMENT NEGATIVE
FIXED POINT CHOPPED	$0 \geq e \geq -M$*	$0 \leq e \leq M$	$0 \geq e \geq -M$
FIXED POINT ROUNDED	$-1/2M \leq e \leq 1/2M$		
FLOATING POINT CHOPPED	$0 \geq e \geq -2^c M$	$0 \leq e \leq 2^c M$	$0 \geq e \geq -2^c M$
FLOATING POINT ROUNDED	$-2^{c-1}M \leq e \leq 2^{c-1}M$		

$*M = 2^{-b_2} - 2^{-b_1}$

TABLE 2.2 TRUNCATION ERROR BOUNDS

2.2.4 Autopilot errors due to parameter inaccuracy

In the design of the digital autopilot and the selection of parameters, attention must be given to the effect of specifying the parameters with a finite word length. This is true both in design by digitizing a continuous controller and by the optimal digital autopilot and observer design. The actual realization of the controller will have parameters chosen from that set of numbers whose binary equivalent can be represented by a specified word length. The process of system design under the imposition of such a constraint, however, is extremely difficult. A more logical approach is to design the system, chosing the parameters without regard to which numbers can be stored with no error, and then rounding the parameter values to fit the registers. Such a design would be suitable, depending upon the sensitivity of critical properties of the system to autopilot parameters. Hence in the early phases of autopilot design by either method, a sensitivity analysis is called for.

The pulse transfer function of a linear, constant coefficient, discrete autopilot can be written

$$H(z) = \frac{\sum_{k=0}^{m} a_k z^k}{\sum_{k=0}^{n} b_k z^k} \qquad (2.18)$$

where bn may be set equal to 1 without loss of generality. The poles of H(z) are related to the coefficients by

$$\prod_{k=1}^{n} (z-z_k) = \sum_{k=0}^{n} b_k z^k \qquad (bn=1) \qquad (2.19)$$

The sensitivity of a pole z_j to a coefficient b_i may be examined via

$$Sz_j = \frac{\partial Z_j}{\partial b_i}\bigg|_{z=z_j} sb_i$$

where

$$\frac{\partial z_j}{\partial b_i}\bigg|_{z=z_j} = \frac{-z_j^{\,i}}{\prod_{\substack{k=1 \\ k\neq j}}^{n} (z_j-z_k} \qquad (2.20)$$

the sensitivity is written

$$\left|\frac{SZ_i}{Z_i}\right| = \left|\frac{|Z_j|^{i-1}|b_i|}{\prod_{\substack{k=1\\k\neq j}}^{n} |Z_j - Z_k|}\right| \cdot \left|\frac{Sb}{B}\right| \qquad (2.21)$$

The sensitivity is most critical for poles in the region $Z_i \quad 1$ near the stability axis. Where the sensitivity may be approximated

$$S = \frac{|b_i|}{\prod_{\substack{k=1\\k\neq j}}^{n} |Z_j - Z_k|} \qquad (2.22)$$

Pulse transfer pole motion due to rounding the coefficients b_k in the implementation of the autopilot is seen to be a function of the following:

(1) Order of the system

(2) Location and distribution of the poles (determined in part by the s to z transformation and the sampling frequency)

(3) Magnitude of the coefficient.

The effect of mapping a simple pole at s=-a to the z plane for several transformations is shown below.

First Order Integration: $z = 1-Ta$

Second Order Integration: $z = 1-Ta+\frac{T^2a^2}{2}$

Convolution/Impulse Invariance: $z = e^{-aT}$

Bilinear: $z = \frac{1+aT/2}{1-aT/2}$

In each case, as the sampling period T is decreased, the pole is mapped near z=1. A general property of these transformations is that as the sampling rate is increased above the Nyguist frequency, the poles of the pulse transfer function are clustered near $Z=1$. The effect is to increase the sensitivity of the poles to the accuracy of the coefficient.

If the coefficients have infinite precision, the stability properties of the system are preserved across most of the transformations. However, for finite word length representations of the coefficients, stability is an important consideration. For small magnitudes of $Z_i - Z_k$ the stability of the system may be considerably effected by roundoff particularly in high order systems.

In cases where it is necessary to realize a design with lightly damped poles and relatively high sampling rates, it is advantageous to separate the autopilot topologically into lower order subsystems. In which case a cascade or parallel realization of the design may be implemented to avoid the pole wandering inherent in high order systems.

3 DISCRETE OPTIMAL CONTROL

3.1 Problem Formulation

This chapter considers the application of discrete linear regulator theory to the design of a digital autopilot. The results are collected below in a simple form. These results, along with complete derivations, may be found in optimal control and estimation texts.

The linear, continuous time, controllable and observable system for the problem is specified in vector form

$$\dot{x} = A(t)x + B(t)u \qquad (3.1)$$

$$y = C(t)x + v \qquad (3.2)$$

The matrices $A(t)$, $B(t)$, and $C(t)$ are the state, control, and measurement coefficient matricies. The vector x is the state vector which incorporates vehicle-target relative motion. The vector u is the control vector derived from the autopilot control law. Since the autopilot is digital, the components of u are piecewise constant. The vector y is the available measurement of the state vector x, corrupted by noise. The noise vector v is zero mean white noise with covariance

$$\text{cov}\,\{v(t),\, v(t)\} = V(t)S(t-\tau) \qquad (3.3)$$

Since the control vector is piecewise constant, the state equations may be discretized with no error. The integrated state equation is

$$x(t) = F(t, to)\ x\ (to) + \int_{to}^{t} F(t,\tau)\ B(\tau)u(\tau)d\tau \qquad (3.4)$$

where $F(t,to)$ is the state transition matrix corresponding to $A(t)$. Since $u(t)$ is piecewise constant over the sampling period

$$x(nT+T) = F(nT+T,nT)\ x\ (nT) + \left[\int_{nT}^{(n+1)T} F(nT+T,\tau)\ B(\tau)d\tau\right] u(n\tau) \qquad (3.5)$$

With a slight change of notation, the discretized state equation becomes exactly (without approximation)

$$x(n+1) = F(n+1,n)\ x\ (n) + G(n+1,n)\ u(n) \qquad (3.6)$$

The stage dependent matrices $F(n+1,n)$ and $G(n+1,n)$ may be computed from $A(t)$ and $B(t)$, although the state transition matrix for a time varying system is often difficult to compute.

The discrete optimal control law for u will be shown for a general quadratic performance index in the state and control. Such a solution will require the implementation of u from the state vector x.

$$u(n) = L(n)\ x\ (n) \qquad (3.7)$$

In the absence of the state vector, the control in the linear, quadratic, gaussian case is the same implementation of the mean of the state vector conditioned on the sequence of measurements.

$$u(n) = L(n)\ E\{x(n)|y(o),y(1)\ .\ .\ .y(n)\}$$

The following section will give a brief development of the optimal control law. The development of the optimal estimate of the state follows in the third section.

3.2 Discrete Optimal Control

Given the system

$$x(n+1) = F(n+1,n)x(n) + G(n+1,n)u(n)$$

and a quadratic performance index

$$J \tfrac{1}{2}x^T(Nf)Q(Nf)x(Nf) + \tfrac{1}{2}\sum_{n=0}^{Nf-1} x^T(n)Q(n)x(n) + u^T(n)R(n)u(n) \qquad (3.9)$$

it is desired to choose the sequence u(0), u(1), . . . u(Nf-1) to minimize J. If the Hamiltonian is formed as

$$H = \tfrac{1}{2}x^T(n)\; Q(n)x(n) + \tfrac{1}{2}u^T(n)\; R(n)u(n) + \lambda^T(n+1)x(n+1) \qquad (3.10)$$

the necessary conditions for a minimum are [14].

$$\frac{\partial H}{\partial u(n)} = 0 = R(n)u(n) + G^T(n+1,n)\lambda(n+1) \qquad (3.11)$$

$$\frac{\partial H}{\partial x(n)} = \lambda(n) = Q(n)x(n) + F^T(n+1,n)\lambda(n+1) \qquad (3.12)$$

$$\frac{\partial H}{\partial \lambda(n+1)} = x(n+1) = F(n+1,n)x(n) + G(n+1,n)u(n) \qquad (3.13)$$

$$x(0) = Xo \qquad (3.14)$$

$$\lambda(Nf) = Q(Nf)x(Nf) \qquad (3.15)$$

The costate vector is linearly related to the state vector

$$\lambda(n) = P(n)x(n). \qquad (3.16)$$

Using this condition, the necessary minimum conditions can be reduced simul-taneously to

$$P(n) = Q(n) + F^T(n+1,n)\; P(n+1)\; \left[I - G\;(n\pm1,n)\; H(n) G\;(n+1,n)\; P(n+1)\right] F(n+1,$$

(3

$$H(n) = \left[R(n) + G^T(n+1,n)\; P(n+1)\; G(n+1,n)\right]^{-1} \qquad (3.18)$$

$$u(n) = -H(n)\; G(n+1,n) P\;(n+1)\; F(n+1,n)x(n) \qquad (3.19)$$

The first two equations are the discrete equivalent of the Ricatte equation, and may be solved off-line in reverse stage for the sequence P(Nf-1), P(Nf-2), . . . P(0) subject to the boundary condition

$$P(Nf) = Q(Nf)$$

The matrix inverse indicated above is replaced by the unique pseudo-inverse in the singular case. When the central weighting matrix

$R(n)$ is positive definite for each n, the equation for $P(n)$ may be simplified

$$P(n) = Q(n) + F^T(n+1,n)\ M(n) \tag{3.20}$$

$$P(Nf) = Q(Nf) \tag{3.21}$$

$$M(n) = \left[P^{-1}(n+1)+ G(n+1,n)\ R^{-1}(n)\ G^T(n+1,n)\right]^{-1} F(n+1,n) \tag{3.22}$$

$$u(n) = - R^{-1}(n)\ G^T(n+1,n)\ M(n)x(n) \tag{3.23}$$

In the case where the state equations and performance weighting coefficients are time invariant, it is desirable to seek a constant optimal feedb ack gain. If the final time is taken large enough that the steady-state gain may be used, the constant Pmatrix is obtained from the algebraic equation

$$P = Q + F^T P \left[I-G\ (R+G^T PG)^{-1} GP\right] F \tag{3.24}$$

and the optimal control is

$$u = - \left[R + G^T PG\right]^{-1} GPF \quad x(n) \tag{3.25}$$

The approximation due to taking an infinite final time is usually good when the actual operating time is large compared to the dominant system time constants.

In the target directed vehicle, the kinematic coefficients are range dependent, hence the time-invariance assumption is usually not good. One approach for this situation is to compute the optimal steady-state feedback gain matrices for the system using the aerodynamic and kinematic coefficients at several different points in the trajectory. These gains can then be switched in the autopilot at the appropriate point in the flight path.

The optimal control requires the state vector for its implementation. When the State vector is not available, which is invariably the case, it, or its best estimate, must be obtained from the measurement vector. In the linear, quadratics gaussian case, the minimum variance estimate of the state vector may be used in lieu of the inaccessible state vector. The procedure for obtaining the estimate is approached in the next section.

3.3 Optimal State Estimation

For the linear, quadratic, gaussian control problem that is formulated in section 3.1, the optimal control law, developed in 3.2, may be applied to the mean of the state vector conditioned on the sequence of measurement vectors [15].

$$u(n) = L(n)\, E\{\, x(n)/y(1),\, y(z),\, .\,.\,.\; y(n)\}$$

The minimum variance estimate (as well as others in this case) is the conditional mean estimate, which may be computed sequentially by a Kolman filter [16, 17]. A heuristic development of this very well known technique is given below. The applicability of this technique in the guidance loop of the autopilot is obvious.

The discrete stochastic control problem of concern is as follows:

$$x(j+1) = F(j+1,j)x(j) + G(j)u(j) \tag{3.26}$$

$$y(j) = H(j)x(j) + v(j) \tag{3.27}$$

$$E\{v(j)\} = 0 \qquad j = 0,1,\;.\;.\;. \tag{3.28}$$

$$E\{v(i)v(j)^T\} \begin{cases} Vv(j) & j=i \\ 0 & \text{otherwise} \end{cases} \tag{3.29}$$

$$E\{x(0)\} = xo \qquad (3.30)$$

$$E\{x(0)\ x^T(0)\} = Vx(0) \qquad (3.31)$$

$$E\{V(j)\ x^T(0)\} = 0 \qquad j=0,1,.\ .\ . \qquad (3.32)$$

Let $\hat{x}\ (j/i)$ represent the estimate of $x(j)$ given the set of measurements $y(o)$, $y(1)$, . . . $y(i)$. The one step propagation of the estimate without additional information is

$$\hat{x}\ (j+1|J) = F(j+1,j)x(J|j) + G(j)u(j) \qquad (3.33)$$

It can be shown that the new linear estimate at $j+1$ is the combination of the propagated estimate and a weighted innovation using the current measurements.

$$\hat{x}(j+1|j+1)=\hat{x}(j+1|j) + K(j+1)\left[y(j+1)-H(j+1)\hat{x}(j+1|j)\right] \qquad (3.34)$$

The error in estimation is denoted

$$\tilde{x}(j|i) = x(j) - \hat{x}(j|i). \qquad (3.35)$$

Then

$$\tilde{x}(j+1|j+1) = \tilde{x}(j+1|j) - K(j+1)\left[H(j+1)\tilde{x}(j+1|j)-v(j+1)\right] \qquad (3.36)$$

$$\tilde{x}(j+1|j) = F(j+1,j)\tilde{x}(j|j) \qquad (3.37)$$

If the covariance matrices are denoted

$$V\tilde{x}(j|i) = E\{\tilde{x}(j|i)\tilde{x}^T(j|i)\} \qquad (3.38)$$

$$Vv(j) = E\{V(j)V^T(j)\} \qquad (3.39)$$

then

$$V\tilde{x}(j+1|j) = F(j+1,j)V\tilde{x}(j|j)F^T(j+1,j) \qquad (3.40)$$

and

$$V\tilde{x}(j+1|j+1) = \left[I-K(j+1)H(j+1)\right]\ V\tilde{x}(j+1|j)\left[I-K(j+1)H(j+1)\right]^T+K(j+1)Vv(j+1)K^T(j+1) \qquad (3.41)$$

Since x(j+1/j) and V)j+1) are uncorrelated. The variance of the estimate error to be minimized, $E\{\tilde{x}^T(j+1|j+1)x(j+1|j+1)\}$, is the trace of the covariance matrix $V\tilde{x}(j+1|j+1)$. Given $V\tilde{x}(j+1,j)$ the variance of the estimate error in the j+1-st step can be minimized by the choice of $K(j+1)=V\tilde{x}(j+1|j)H^T(j+1)\left[Vv(j+1)+H(j+1)V\tilde{x}(j+1|j)H^T(j+1)\right]^{-1}$

$$\tag{3.42}$$

This choice of K(j+1) may be used to simplify the expression for $V\tilde{x}(j+1|j+1)$ which becomes

$$V\tilde{x}(j+1|j+1) = \left[I-K(j+1)H(j+1)\right] V\tilde{x}(j+1|j). \tag{3.43}$$

The equations for K(j+1), $V\tilde{x}(j+1|j)$, and $V\tilde{x}(j+1|j+1)$ may be solved sequentially off-line and the sequence of K(j)'s stored for on-line generation of the optimal estimate. The Kalmon filter equations are repeated in Table 3.I for convenience.

One Step Error Variance	$V\tilde{x}(j+1|j) = F(j+1,j)V\tilde{x}(j\surd j)F^T(j+1,j)$
Kalman Gain	$K(j+1) = V\tilde{x}(j+1|j)H^T(j+1)\left[Vv(j+1)=H(j+1)V\tilde{x}(j+1|j)H^T(j+1)\right]^{-1}$
Corrected Error Variance	$V\tilde{x}(j+1|j+1) = \left[I-K(j+1)H(j+1)\right] V\tilde{x}(j+1|j\ )$
One Step Error Estimate	$\hat{x}(j+1|j) = F(j+1,j)\hat{x}(j|j)+G(j)u(j)$
Corrected Error Estimate	$\hat{x}(j+1|j+1) = \hat{x}(j+1|j)+K(j+1)\left[y(j+1)-H(j+1)\hat{x}(j+1\ j)\right]$

TABLE 3.1 KALMAN FILTER EQUATIONS

As in the case of the linear optimal control it is tempting to seek a reduction in the amount of storage and computation required by the Kalman filter. However, instability or divergence of the estimation error, sometimes a problem in the best of cases, becomes a severe restriction under reduction techniques.

It can be shown that, if the linear system is completely state *observable*, the estimate error process is stable and an estimate of the state is available regardless of the initial assumption of prior statistics of the state. In the actual case, however, the system model used in the Kalman filter algorithms is inaccurate and the statistics of the measurement noise are in error. Hence, divergence of the filter must be considered.

1. Delansky, J., "A Common Review of Frequency Domain Theory for Continuous and Discrete Time Linear Time Invariant Systems", AFATL TR (In Preparation), USAF, 1975.

2. Gold, B. and Rader, C.M., Digital Processing of Signals, McGraw-Hill, NY, 1969.

3. Oppenheim, A.V. and Schafer, R.W., Digital Signal Processing, Prentice-Hall, NJ, 1975.

4. Kaiser, J.F. and Kuo, F.F., System Analysis by Digital Computer, Wiley, NY, 1966.

5. Kaiser, J.R., "Design Methods for Sampled Data Filters", Proc. 1st Allerton Conf. Circuit System Theory, November 1963, pp. 221-236.

6. Rader, C.M. and Gold, B., "Digital Filter Design in the Frequency Domain", Proc. IEEE , Vol. 55, February 1967, pp. 149-171.

7. Golden, R.M. and Kaiser, J.F.,"Design of Wideband Sampled-Data Filter", Bell System Tech. J., Vol. 43, No. 4, Pt. 2, July 1964, pp. 1533-1545.

8. Lue, B. and Kaneko, T., "Error Analysis of Digital Filter Realized with Floating-Point Arithmetic", Proc. IEEE, Vol. 57, October 1969, pp. 1735-1747.

9. Sandberg, I.W., "Floating-Point-Roundoff Accumulation in Digital Filter Realization", Bell System Tech. J., Vol. 46, October 1967, pp. 1775-1791.

10. Lue, B., "Effect of Finite Word Length on the Accuracy of Digital Filters - A Review", IEEE Trans. Circuit Theory, Vol. CT-18, November 1971, pp. 670-677.

11. Oppenheim, A.V. and Weinstein, C.J., "Effects of Finite Register Length in Digital Filtering and the Fast Fourier Transform", Proc. IEEE, Aug. 1972, pp. 957-976.

12. Kaneko, T. and Liu, B., "Accumulation of Roundoff Error in Fast Fourier Transforms", J. Assoc. Comput. Mach., Vol. 17, Oct. 1970, pp. 637-654.

13. Chan, D.S.K. and Rabiner, L.R., "Theory of Roundoff Noise in Cascade Realizations of Finite Impulse Response Digital Filters", Bell System Tech. J., Vol. 52, No. 3, Mar 1973, pp. 329-345.

14. Fan, L.T. and Wong, C.S., The Discrete Maximum Principle, Wiley, NY, 1964.

15. Joseph, P.D. and Tou, J.T., "On Linear Control Theory", AIEE Trans, Vol. 80, pp. 193-196, 1961.

16. Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems", Trans. ASME, J. Basic Eng., Vol. 82D, pp. 34-45, Mar. 1960.

17. Kalman, R.E. and Bucy R., "New Results in Linear Filtering and Prediction Theory", Trans. ASME, J. Basic Eng., Vol. 83D, pp. 95-108, Mar. 1961.

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
EGLIN AFB, FLORDIA

(CONDUCTED BY AUBURN UNIVERSITY)

OHMIC CONTACTS

FOR TRANSFERRED ELECTRON DEVICES

Prepared by:	Bruce P. Johnson PhD.
Academic Rank:	Associate Professor
Department and University:	Department of Electrical Engineering University of Nevada
Assignment:	
(Laboratory)	Avionics
(Division)	Electronic Technology
(Branch)	Microwave Technology
USAF Research Colleagues:	G. L. McCoy/C. I. Huang PhD.
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

OHMIC CONTACTS FOR TRANSFERRED ELECTRON DEVICES

By

Bruce P. Johnson

ABSTRACT

Gallium arsenide transferred electron devices (TED's) have a promising future in microwave integrated circuits. They can be used in threshold logic as well as in frequency division and multiplication and other signal processing applications where GHz frequencies are required.

One of the current challenges facing the technology is to make reliable low resistance ohmic contacts to these devices. This report presents an evaluation of the current status of the contacting technology which shows that the current process used in an in-house TED program is state-of-the-art but that state-of-the-art is not good enough to keep the contact resistance at 1-10% of the total device resistance. The report presents the results-to-date of forming n+ layers by shallow donor diffusion to lower specific contact resistance. The results indicate that lowered contact resistance can be achieved by this technique but an additional order of magnitude improvement will require the understanding and control of a thin surface film formed at high sulfur to arsenic ratios and/or at high diffusion temperatures.

Acknowledgements

The author gratefully acknowledges Chief Gary McCoy for suggesting the problem and for making this Summer Fellow feel right at home in his group. Special thanks go to Dr. Chern Huang for many valuable discussions and assistance, Lt. Robert Lyons for finding necessary equipment at TER and providing ion implanted samples, Captain Ted Lane for providing epitaxial GaAs wafers and Charles Cox for assistance in sample processing.

Organization

INTRODUCTION

The transferred-electron effect is one of the most important and interesting new effects in semiconductors. Transferred-electron devices (TED's) have enormous potential for signal processing at microwave frequencies. Before the full capability of this new technology can be realized, it will be necessary to develop a planar technology with well defined and controlled material parameters.

One of the most critical of these parameters is the metal semiconductor interface. For three terminal TED's, it is necessary to be able to make ohmic contacts with low contact resistance (1-10% of total device resistance) as well as Schottky barrier gates with well defined stable barrier height. In the ten week period of this program, the author has investigated the ohmic contact problem to GaAs TED's as part of an in-house research program on these devices.

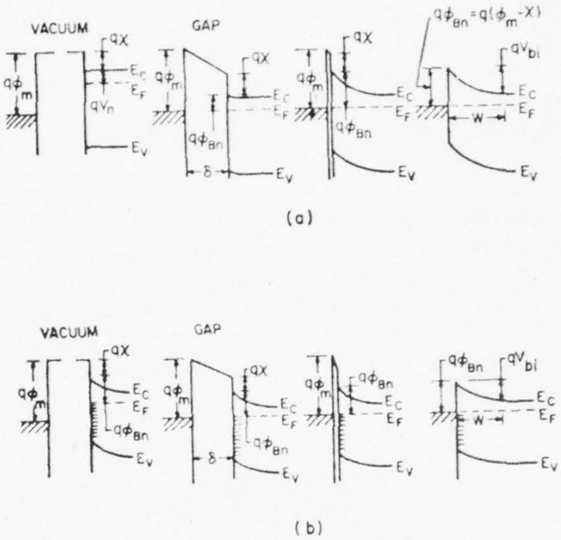
THE GENERAL PROBLEM

Figure 1 shows the electronic energy relations between a metal and an n-type semiconductor for the two extreme cases of no surface states (1a) and sufficient surface states to pin the Fermi level so that the barrier height is independent of the metal work function (1b)[1]. At far left, the metal and semiconductor are not in contact and the system is not in thermal equilibrium. At the far right they are in contact and the system is in thermal equilibrium with no interfacial layer. Current flow across the metal-semiconductor interface is due to thermionic emission (or diffusion), thermionic-field emission, or pure field emission depending on barrier height and width. "Ohmic" contacts require barriers such that field or thermionic-field emission can take place. There are four accepted approaches to controlling barrier height and width in obtaining ohmic metal-semiconductor junctions. All four will be briefly discussed even though this study is concerned only with the first.

1. n^+ LAYER - If the semiconductor is weakly doped then a sizable barrier will exist due to the Fermi level-conduction band distance and the width of the charge depletion region. This is the case with GaAs TED's where device requirements place the epitaxy doping at 10^{15}-10^{16} donors/cm^3. The barrier can be both lowered and narrowed by forming an n^+ layer near the surface of the semiconductor[2]. There are three common processes for forming an n^+ layer, namely ion implantation, higher doping during the last stage of epitaxy growth, and diffusion. All are limited by the solubility of donors in GaAs which is in the mid 10^{18}/cm^3(2). Ion implantation is currently being investigated by Lt. Robert Lyons of this Laboratory. The major questions to be answered for GaAs are how high a doping level can be implanted and how well can implantation damage be annealed out. A recent study at Air Force Cambridge Research Laboratories(3) indicated that up to 10^{18}/cm^3 electrical active donors can be implanted in GaAs under suitably controlled conditions (40 percent

[1] Image force barrier lowering has been neglected.

[2] The required thickness is not well defined as some alloying processes diffuse the contacting species up to 2.0 μm. The purpose of the n^+ layer is to decrease the depletion width from ~4000 A to ≤100 A.



ϕ_m = Work function of metal

χ = Electron Affinity of Semiconductor

ϕ_{Bn} = Barrier Height of Metal-Semiconductor Barrier

V_{bi} = Built in Potential

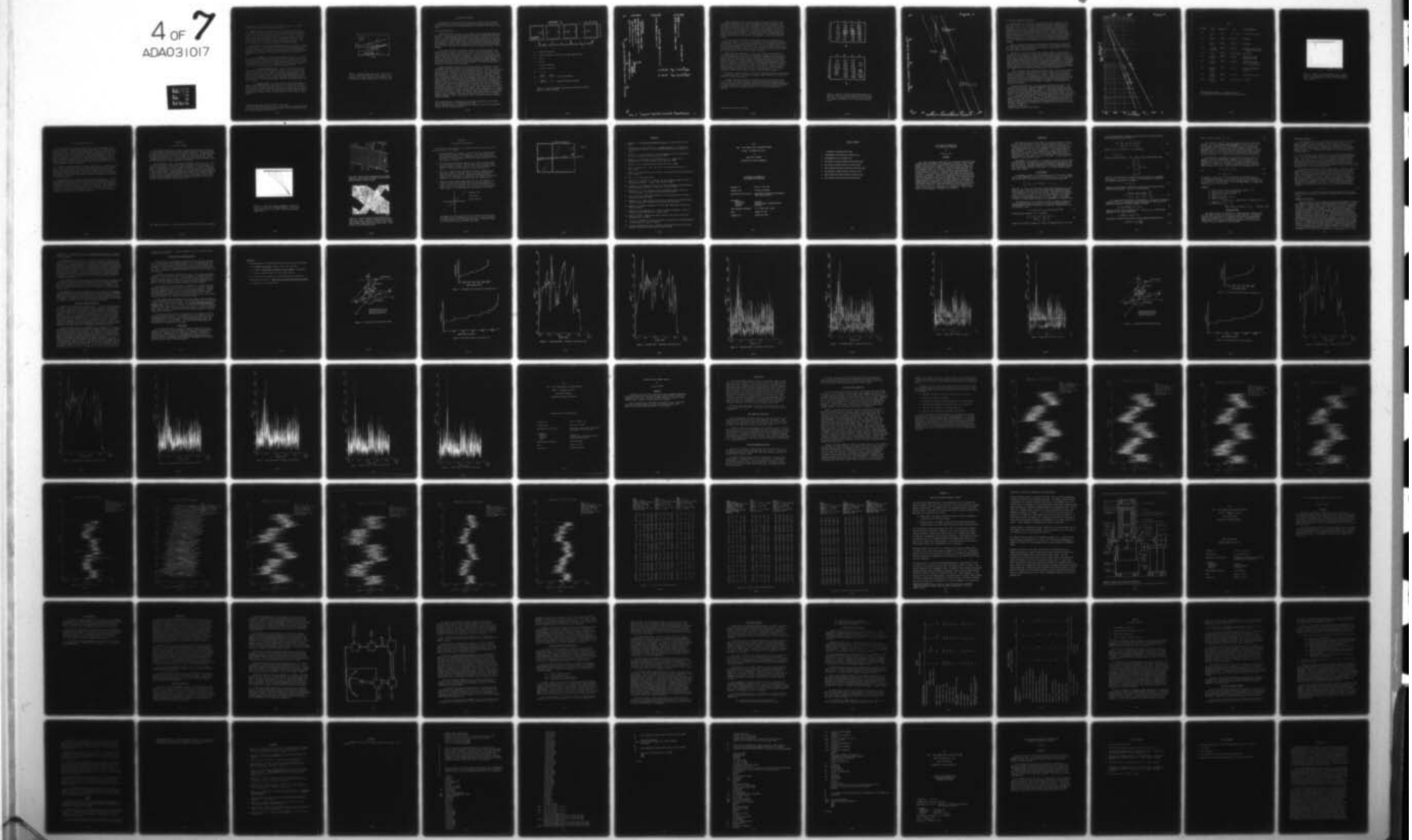
δ = Thickness of Interfacial Layer

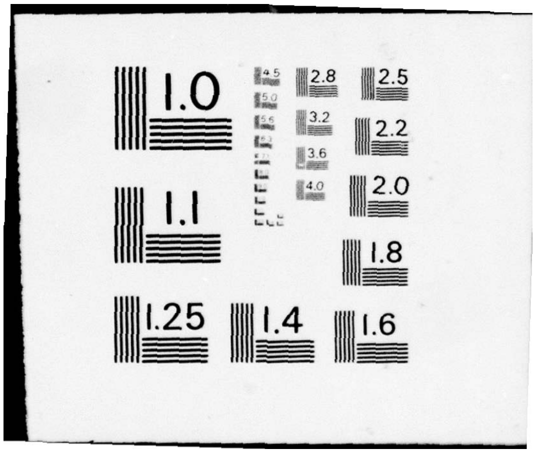
W = Depletion Region Width

Figure 1. Energy band diagrams of metal-semiconductor contacts (a) with and (b) without surface states. (After Ref. 1)

AD-A031 017 AUBURN UNIV ALA SCHOOL OF ENGINEERING F/G 5/2
1975 USAF/ASEE SUMMER FACULTY RESEARCH PROGRAM.(U)
SEP 75 J F O'BRIEN F44620-75-C-0031
UNCLASSIFIED AFOSR-TR-76-1113 NL

4 OF 7
ADA031017




1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

utilization of the implanted ions). No attempt was made in the study to make ohmic contacts to the implanted material.

Higher doping in the last stage of the vapor phase epitaxy growth is a second way to form an n^+ layer. This technique has not been specifically tried on TED's although it has been used in other III-V materia devices. The question to be answered for TED's is how sharp a gradient in the doping profile can be obtained. Since masking during this last growth stage would be difficult, it will probably be necessary to mesa etch for isolation which would not give a true planar structure. This will increase some of the photolithography resolution problems.

During the ten weeks of this program the author has studied the th approach, diffusion. Although donor diffusion into GaAs has been repor by several authors [4-7], conflicting results have been obtained and nc attempt has been made to measure contact resistance of the diffused lay This approach to forming an n^+ layer will be discussed in more detail below.

2. SMALLER Fg. Narrowing the forbidden band gap in a thin l near the surface accomplishes the same net result as forming an $n+$ laye It is usually more difficult to accomplish by an alloying process but h been successfully performed during vapor growth of epitaxy layers. Ple is believed to be using the alloy approach by forming an ohmic contact with indium[3] followed by the usual gold-germanium nickel (8).

3. LOWER WORK FUNCTION METAL. Figure 1 indicated that the b rier height is directly proportional to the metal work function when nc surface states are present. Figure 2 shows the variation of the barrie height with metal work function for four different semiconductors. GaA does not show the wide variation in barrier height with metal work function that is characteristic of silicon. Hence the metal work function not expected to be an important variable parameter in obtaining ohmic c tacts to GaAs.

4. SURFACE STATES. The reason that the barrier height does depend strongly on metal work function is that the semiconductor Fermi level is partially pinned by surface states on the GaAs as was illustra in Figure 1b. Very little work has been done in understanding the natu of the surface states on GaAs. This would appear to be a fruitful area investigation as surface states are also expected to affect device perf mance[4].

[3] InAs has a band gap of 0.33 eV verses 1.43 for GaAs.

[4] The author has observed several surface state effects in gallium phos such as surface treatment before metallization, effective passivation, p-n junction leakage current effects.

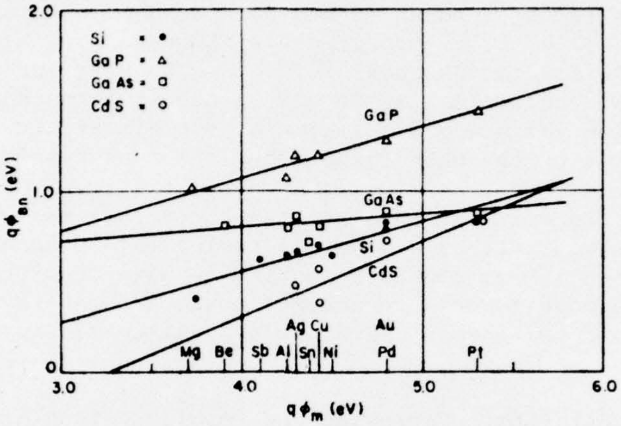


Figure 2. Barrier height verses metal work function for four semiconductor materials (9). Gold-12 wt % germanium, the contact material most frequently used on GaAs gives a barrier height of 0.6 to 0.7 eV (10).

The Specific Problem

During the ten week period of this study, the author has concentrated on characterizing the current contacting process and in shallow diffusing sulfur into GaAs to form an n+ layer and measuring the resulting specific contact resistance.

A. The Current Process

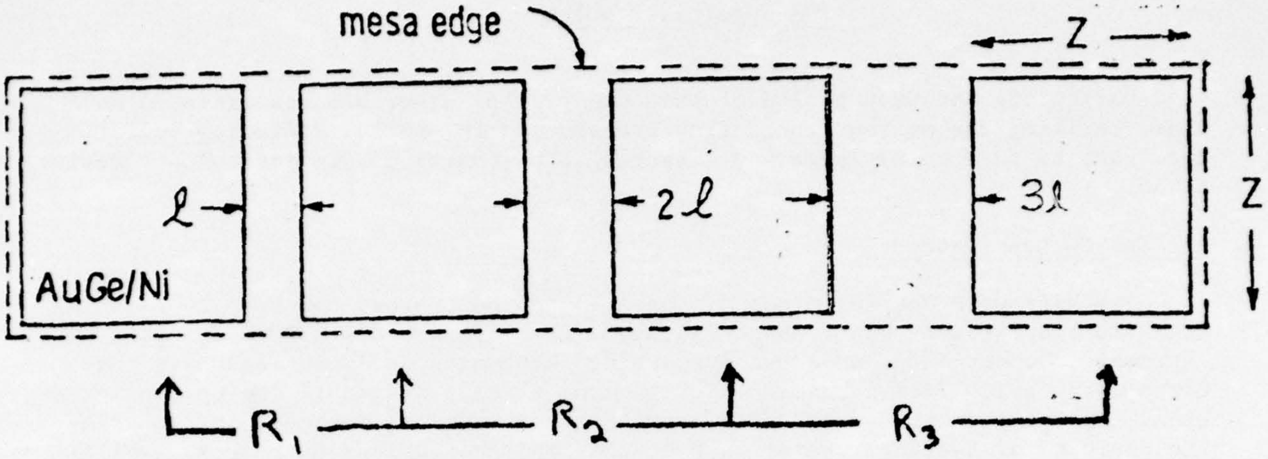
The Microwave Sources Group of the Avionics Laboratory has been fabricating GaAs TED's by growing vapor phase epitaxy (VPE) layers of 10^{15}-$10^{16}/cm^3$ GaAs on external vendor chromium doped insulating substrates, evaporating ohmic contacts using gold-12wt% germanium-nickel, mesa etching desirable device configurations, applying an evaporated gold gate on three terminal devices, and mounting the devices for testing. While some of the finished devices have shown typical TED performance, there has been considerable variation in parameters with some devices from the same wafer and some whole wafers not showing acceptable TED performance.

The doping uniformity and thickness control of the VPE process has been checked and this process is believed to be well controlled and uniform. Each VPE is checked for doping level and thickness. The next processing step is the vacuum evaporation of gold-12wt% germanium followed by a second evaporation with nickel[5] to prevent "balling" of the contact. Evaluation of the metallization process had just been initiated when the author started the Summer Program. Figure 3 shows the contact resistance mask and associated measurement parameters required to measure the specific contact resistance. The mask and procedure was developed by C. Huang of this Laboratory.

Figure 4 presents typical specific contact resistance ranges for two typical samples which were subjected to varying metallization processing. Sample C23.18A, B, and C were segments from the same VPE process. A and B were evaporated by the standard procedure at the same time while C had sputtered gold-12wt% germanium contacts[6] followed by an evaporated nickel overlay. The three samples were annealed by the standard heat cycle process (heated in hydrogen up to 475°C for a few seconds), measured, and then subjected to varying heat cycles as indicated. Sample C23.18A was the first sample measured and initial specific contact resistance was 1.4-1.6 X $10^{-2}\Omega$-cm^2 on two randomly chosen mesas. Measurements made after this were on the same mesas chosen to represent a typical range on the sample. Sample C23.18A was measured after 6 minutes, 9 minutes, and 18 minutes in hydrogen at 475°C. It was then sealed in a quartz tube (~$7cm^3$) at mid 10^{-6} torr with 60 mg As and heated at 500°C for 21 minutes. Upon removal from the furnace the tube was quenched in water to prevent arsenic vapor condensation on the sample. As is seen from figure 3, samples C23.18A and B were both improved by an As anneal while C15.15 was slightly worse. The arsenic anneal should assist in preventing arsenic vacancy formation (acceptors) and yet allow the generation of gallium vacancies (donors). Hence at low doping levels, the specific contact resistance should be lowered.

[5] The evaporation boat is changed before the nickel evaporation so the sample sets in room ambient for a varying period of time.

[6] Contact thickness was not measured but the sputtered contacts were considerably thicker than the evaporated contacts.



R_i = measured resistance

r_p = probe resistance, both probes on same evaporated pad

l = 50 μm

Z = 200 μm

α = epitaxy thickness

ρ = epitaxy resistivity

R_i' = $R_i - r_p$

R_c = $\frac{2R_1' - R_2'}{2}$ = $\frac{3R_1' - R_3'}{4}$ = Contact Resistance

r_c = $\frac{(ZR_c)^2\alpha}{\rho}$ $\Omega - cm^2$ = Specific Contact Resistance

Figure 3. Contact resistance mask and formulas used to measure specific contact resistance.

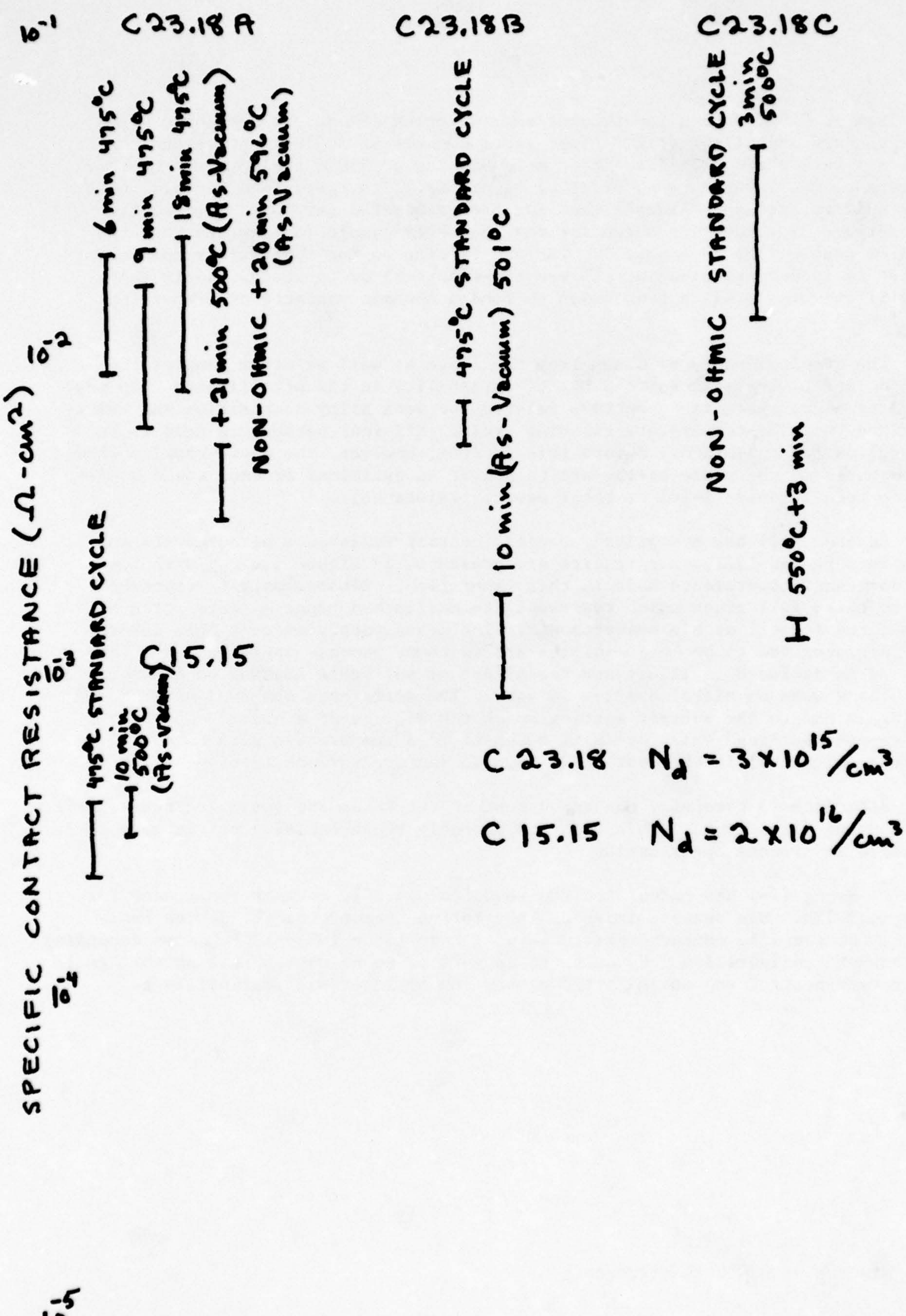


Fig. 4 Typical Specific Contact Resistances

Sample C23.18C with the thicker metallization did not become ohmic with the standard annealing cycle. After three minutes at 500°C in hydrogen the contacts were ohmic and after three more minutes at 550°C the specific contact resistance was lowered by an order of magnitude. The resistance was also much more uniform across the sample than for the evaporated samples. Typical R_1 (see Figure 3) values are shown for the sputtered sample (5a) verses the evaporated sample (5b) in Figure 5. The pad lifting on the sputtered sample may be due to inadequate cleaning (15 sec back sputter) or to the unusually thick metallization. It was not uncommon to find non-ohmic contacts on evaporated samples.

The conclusions to be drawn from the above as well as other samples processed is that there is quite a bit of variability in the metallization process. In particular, there is an optimum relation between alloy composition and thickness and the time-temperature alloying cycle. All four parameters need to be controlled and optimized. Before this is done, however, the above results should be compared to the state-of-the-art to see if an optimized process would satisfy device requirements (1-10% of total device resistance).

Edwards (11) has summarized specific contact resistance measurements to GaAs through mid-1971. His results are presented in Figure 6 along with some new data, and measurements made in this laboratory. Edwards' data is represented by the three full lines which represent the median and range of data which he summarized as well as his measurements. The measurements made in this laboratory are observed to be "state-of-the-art" without process optimization. Reference 12 is included to illustrate the effect of substrate heating on contact resistance when no nickel overlay is used. The difference observed at 10^{16}/cm^3 doping is due to the surface wetting which takes place on a heated substrate verses the "balling" which tends to occur at room temperature with no nickel. The role of nickel in the contact is thus to improve surface wetting.

Reference 13 carefully monitored contact thickness and optimized the annealing cycle using fixed time[7]. This point is probably representative of the gain to be realized in process optimization.

C. Huang (14) has calculated the required specific contact resistance for a typical TED. His results indicate that for one percent of the device resistance, the specific contact resistance will have to be 10^{-15}-10^{-6}ohm-cm^2 depending on contact configuration. Hence there appears to be no way to realize this goal by process control and optimization alone. An n+ layer will definitely be required.

[7] Two minutes at 600°C in nitrogen.

	A	B	C	D
1	180	Pad Lifted	185	Pad Lifted
2	222	324	264	Pad Lifted
3	228	Pad Lifted	Pad Lifted	380
4	226	Pad Lifted	Pad Lifted	Pad Lifted
5	192	210	226	294
6	215	190	198	288
7	234	204	218	255

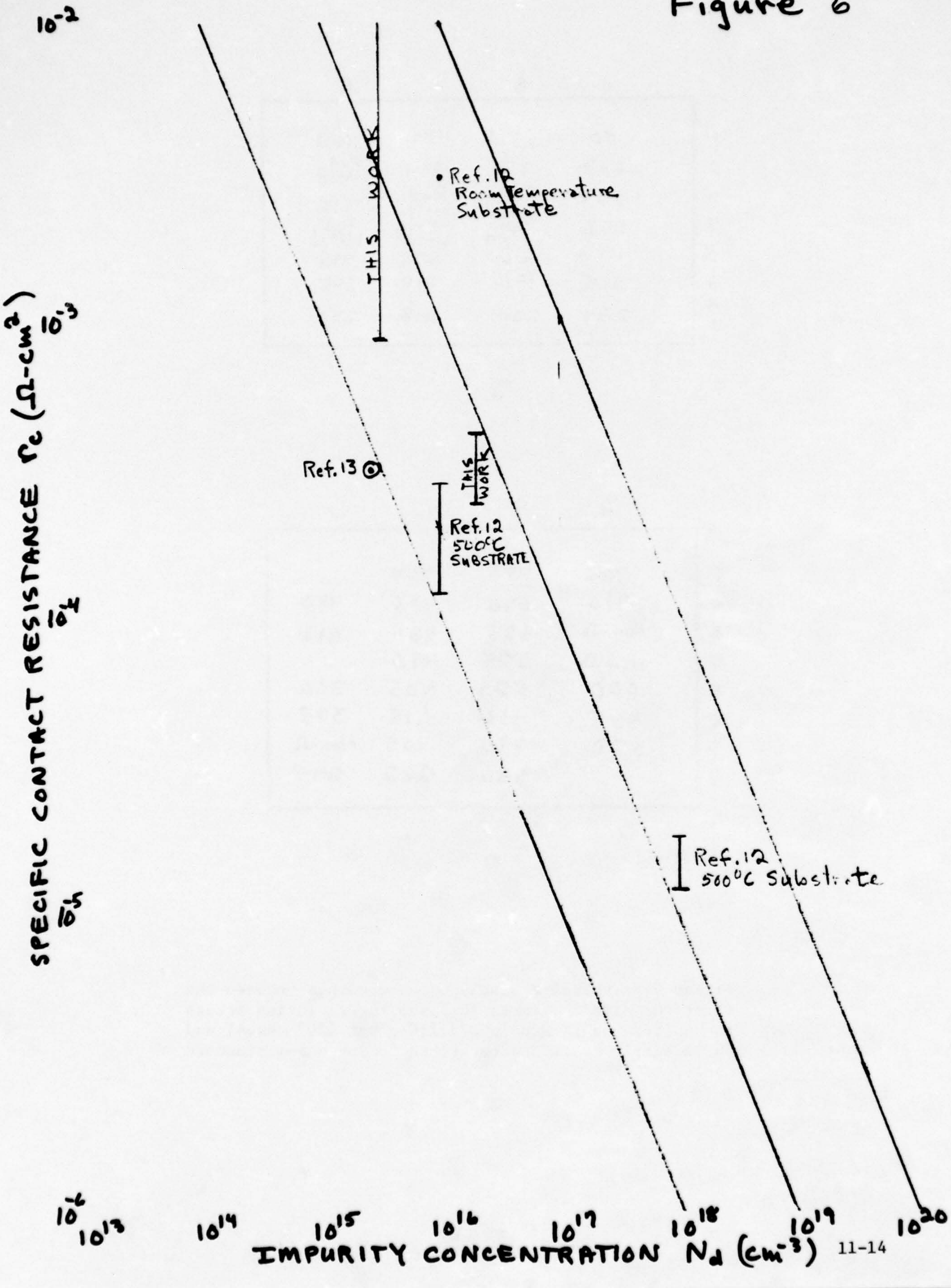
a

	A	B	C	D
1	775	472	454	
2	910	690	435	480
3	Non-Ω	625	436	418
4	Non-Ω	635	410	
5	660	524	405	366
6	485	416	418	375
7	320	290	405	Non-Ω
8		286	282	295

b

Figure 5. Typical resistance measurements between the first two pads on a mesa (R_1) showing variation across the wafer. 5a is sample C23.18C after 550° anneal and 5b is sample C23.18B after first fusing using standard cycle.

Figure 6


10^-2
10^-3
10^-4
10^-5
10^-6
SPECIFIC CONTACT RESISTANCE rc (Ω-cm²)
THIS WORK
• Ref. 12
Room Temperature
Substrate
Ref. 13 ⊙
THIS WORK
Ref. 12
500°C
SUBSTRATE
Ref. 12
500°C Substrate
10^13
10^14
10^15
10^16
10^17
10^18
10^19
10^20
IMPURITY CONCENTRATION Nd (cm⁻³)

B. N+ Layer Formation by Diffusion

Donor diffusion into GaAs has been reported by several authors (15 (Summary through 1972), 16). There are several discrepancies in diffusion coefficient values as well as problems reported with surface compound formation. Germanium, silicon, sulfur, selenium, and tellurium were all considered as candidates for this study. Germanium and silicon were eliminated because of reported difficulty with amphoteric behavior control. Sulfur, selenium, and tellurium have quite similar behavior with all three reported to form surface gallium compounds when the partial constituent pressure is sufficiently high(15). Sulfur was chosen for this study because it has the highest diffusion coefficient of the three and it is being used in the ion implantation experiments of Lt. Lyons. Reference 16 also indicated that sulfur should be relatively well behaved.

Figure 7 presents the temperature dependence of the diffusion coefficient of sulfur in GaAs for GaS:As ratios of 1:2 (and 2:1) and 10:1 after reference 16. Also shown by the dashed line are the results of this study for a 5:1 ratio with data points at 800°C and 900°C.

Ga_2S_2,[8] A_s[9] , and the source wafer were vacuum sealed in a quartz tube at mid 10^{-6} torr and diffused in a three zone Lindberg furnace with a flat zone of <±0.5°C for 12 cm. Typical enclosed volumes were $7cm^3$ and the wafer was separated from the Ga_2S_2 and As by a quartz boat. To prevent deposition of the vapor on the sample surface during cooling the quartz walls were quenched in water upon removal from the furnace while still at the diffusion temperature. A 5:1 Ga_2S_2:As ratio was chosen to minimize any surface compound formation problems yet maintain a high surface sulfur concentration. Reference 16 reported no surface compound problems during Van der Pauw measurements on the diffused sample although samples were "washed in HCl to remove any condensed sulfur from the surface."

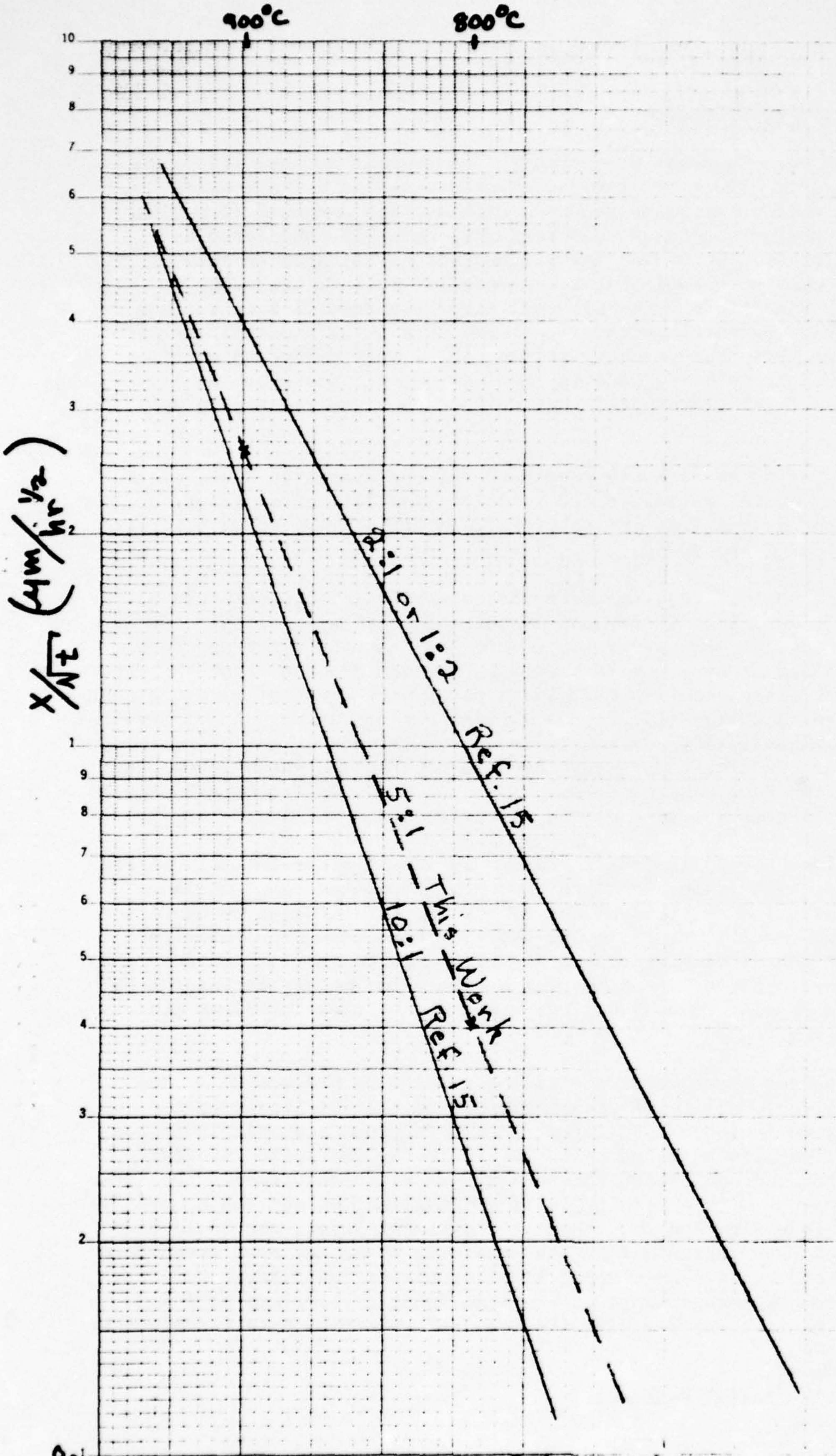
Table I summarizes the diffusions performed to date. After initial calibration, runs were made at 800°C and 900°C to give 2 μm depth (to cover maximum reported alloy depth). The analysis of the effect of the diffusion on specific contact resistance was complicated by two factors. There were problems with a thin surface film on the as diffused material which gave rectifying contacts and the vacuum system developed leak problems which caused the metallization not to stick or gave very poor contact pad quality. Symptoms of these problems are discussed in Appendix A.

Despite the problems there were encouraging results on D-2 and D-7. The analysis of the specific contact resistance for D-2 is made difficult by the fact that the substrate is doped to $\sim 10^{18} cm^{-3}$. A best guess analysis indicates at least an order of magnitude improvement. D-7 (C18.18) was split into three pieces after diffusion and one piece processed at two different times. The first time $r_c = 1\text{-}2 \times 10^{-4} \Omega\text{-}cm^2$ with partially peeling metallization and the second time $r_c \simeq 5 \times 10^{-5} \Omega\text{-}cm^2$. (See Figure 8). Without the diffused layer, one would have expected mid $10^{-3} \Omega\text{-}cm^2$ for this sample. The remaining piece has been given to C. Huang to process into devices. The doping level of the diffused layer for this run was estimated from C_o measurements on aluminum Schottky barriers as 6-7 X $10^{17} cm^{-3}$. This doping level agrees with the measured specific contact resistance according to Figure 6.

[8] Alusuisse 99.9999 Ga_2S_2

[9] Fisher Arsenic Metal, Crystal-Purified.

Figure 7

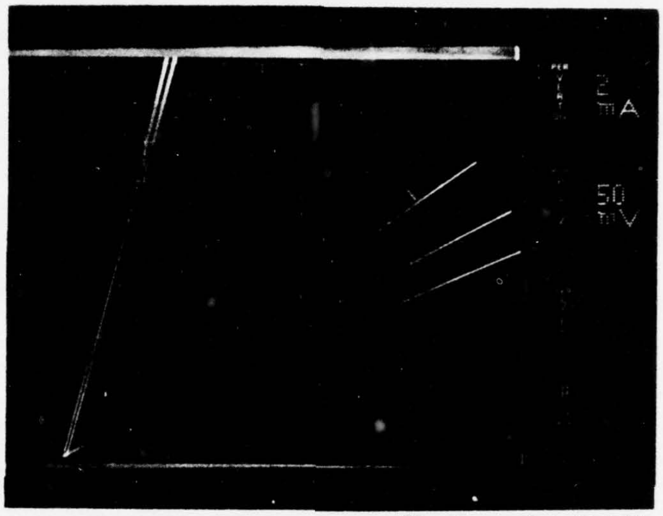


900°C
800°C
10
9
8
7
6
5
4
3
2
1
9
8
7
6
5
4
3
2
0.1
.80
.85
.90
.95
1.0
1.05
1.1
1000/T(oK)
X/√t (μm/hr^1/2)
2:1 or 1:2 Ref. 15
5:1 This Work
10:1 Ref. 15

TABLE I

Diffusion	Sample	Temperature	Time	Results/Comments
D-1	8.11Si C3A	900°C	3 hrs 12 min	4.6μm Diffusion Depth
D-2	8.11Si C3A	898°C	35.5 min	2μm Depth
D-3	C8.07 C3A	800°C	6 hours	1μm Depth
D-4	C23.18A & B, C3A	802°C	25 hours	2μm/Patterned SiO_2 and Sl_3N_4/Film Problems
D-5	C22.11	800°C	24 hrs 30 min	Rectifying
D-6	C18.18 C9.09	898°C	35.5 min	C9.09 Rectifying** C18.18 Rectifying
D-7	C18.18 C9.09	798°C	25 hours	C9.09 Rectifying** C18.18 1-2 X 10^{-4} -cm^2 5 X 10^{-5} -cm^2
D-8	C23.08 T16.15 C23.07	898°C	35.5 min	Capsule Leaked, Wafers Oxidized
D-9	C23.07 C23.08	898°C	35.5 min	10:1 Ratio/In Process
D-10	C23.07 C23.08	800°C	23 hrs 26 min	In Process

*Doped substrate makes r_c calculation uncertain.
**A non-diffused control from C9.09 was also rectifying.



D-7 C18.18

Figure 8. Sample D-7 (C18.18) showing four r_p values and R_1, R_2, and R_3 specific contact resistance measurements. The calculation of r_c gives ~5 X $10^{-5}\Omega$-cm for this measurement.

Conclusions/Recommendations

The limiting factor in lowering the specific contact resistance with a sulfur diffused n+ layer appears to be the electrically active donor concentration at the surface. There are two techniques to raise this concentration to the solubility limit. One is to raise the diffusion temperature and the second is to raise the sulfur to arsenic ratio. Both of these approaches appear to enhance the formation of a surface film as discussed in Appendix A. Hence it will be necessary to determine how to remove this film such that the surface donor concentration remains high and surface states play a minimum role. An alternative approach might be to diffuse through a thin glass film so that this surface compound can not be formed (17).

In order to diffuse n+ layers for device fabrication it is desirable to be able to mask the regions with SiO_2 or Si_3N_4 of known composition and thickness. The one experiment tried in this study (D-4) indicated that further work is definitely required in this area. The pyrolytic reactor which Lt. Lyons is constructing should be an important asset to this problem as pyrolyticly deposited glasses are reported to be better controlled in composition and thickness than sputtered glasses.

The evaporation process should also be improved in three ways. First the gold-germanium alloy and the nickel thicknesses should be monitored. Second, the evaporation should have two boats for sequential evaporation without the intermediate room ambient exposure. Third, the evaporation system should have capability for substrate heating and temperature monitoring. This should help in contact wetting (12), surface state control, and Schottky barrier diode quality (18). With the above evaporation control, the contact annealing cycle can them be optimized for a given metal thickness.

Appendix A

Processing Problems

The evaporation problem was initially encountered during the processing of D-6 and D-7 where some partial and some complete contact pad lifting took place. A non-diffused control piece was then evaluated from T16.15. The pad appearance was quite irregular and the probe resistance (r_p) quite high (44-100 Ω verses 8Ω typical) and variable as is shown in Figure A1. Metallization pads lifted on samples D-9 and D-10 and so the evaporation system is being overhauled. These diffused samples will then be processed a second time.

The film problem is best illustrated by comparing figures A2 and A3. A2 is the surface appearance between mesas (where the mesas are ~3 μm in height) for a non-diffused sample. A3 shows the surface appearance of a wafer diffused at 898°C in a $10Ga_2S_2:1As$ ratio. An attempt to etch 3 μm mesas left very irregular height structures where the white regions are ~3μm deep and the black regions are original surface height. This indicates the presence of an irregular film which is not attacked by the $3HgSO_4:1H_2O_2:1H_2O$ etch. In this figure the metallized specific contact resistance pads have also lifted on the mesas. Before trying to reprocess these samples, they will be recleaned in HCl (15)[8] and in CS_2 to attempt to remove any sulfur type surface film.

[8] The samples were given a 1 minute cleaning in 37% HCl before initial processing.

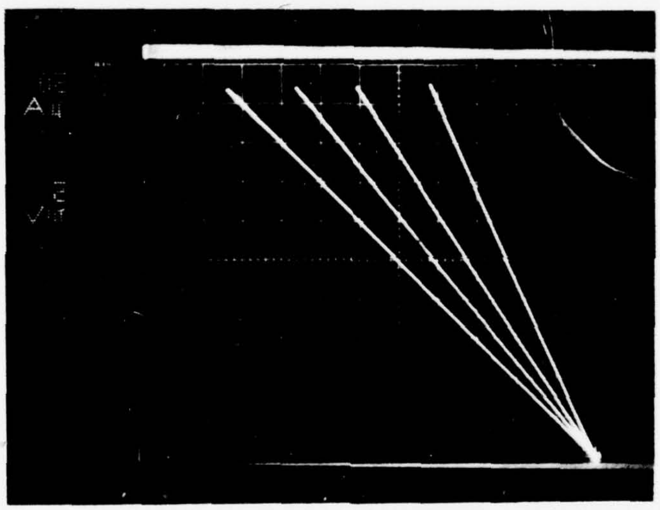


Figure A1. Variation in probe resistance r_p among four pads on one mesa. The 44-100Ω resistance is to be compared with a typical r_p of 6-8Ω. Control non-diffused sample T16.15.

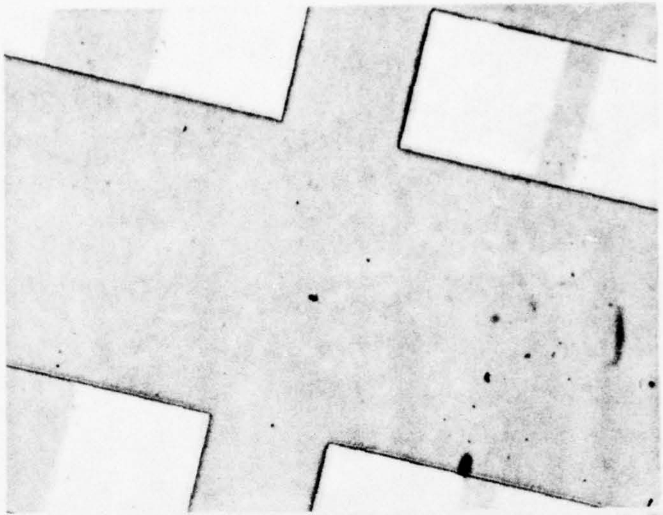


Figure A2. Normal surface appearance after $3H_2SO_4$: $1H_2O_2$:$1H_2O$ etch to give 3 μm mesas with metallized specific contact resistance pads.

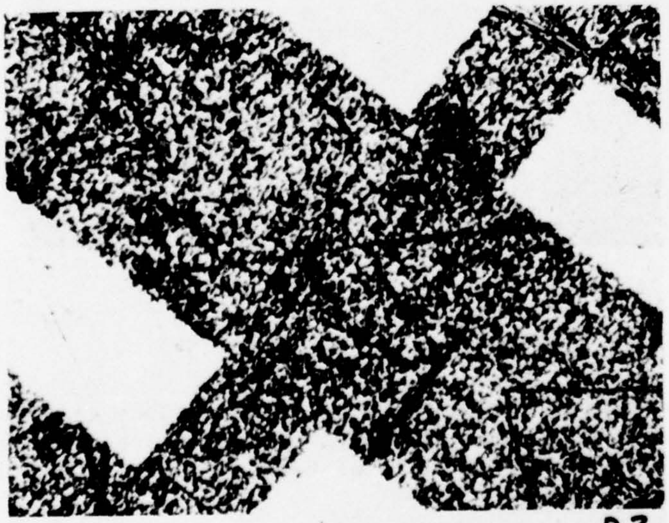


Figure A3. Surface appearance after $3H_2SO_4$:$1H_2O_2$: $1H_2O$ etch to give 3 μm mesa. Metallized pads on the mesas (large white areas) have lifted on this sample. Dark regions between mesas are at mesa height. White areas are ~3μm below mesa level.

Appendix B

Miscellaneous Observations

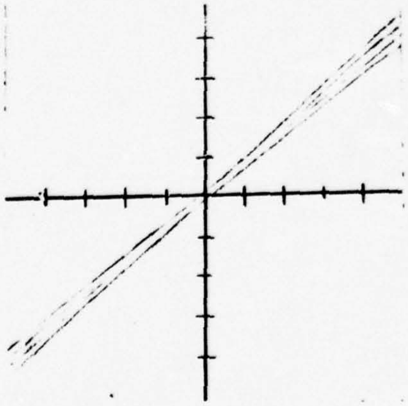
The purpose of this appendix is to document various observations not discussed in the body of the report.

1. During sputtering of the Si_3N_4, a glass slide used to monitor deposition thickness measured 200 A deposition. The Si_3N_4 thickness on the GaAs sample measured on a cleaved edge under the microscope was 30-35,000 A. The SiO_2 showed 220-230A on the glass slide but was also much thicker on the GaAs.

2. The thick-patterned SiO_2 and Si_3N_4 did not crack on the GaAs during the 25 hour 800°C diffusion. The resulting contact resistance measurements showed rectifying contacts. Part of this problem is attributed to a thin glass layer remaining in the patterned areas.

3. C2Z23.09, a Si_3N_4 covered sample which had been ion implanted, did show Si_3N_4 cracking on the GaAs during a 10 minute 800°C arsenic anneal (vacuum sealed capsule). The cracked Si_3N_4 could not be removed by HF (~100 hours) or buffered HF. This suggests the possible formation of an arsenosilicate glass during the arsenic anneal.

4. C2X23.03, a Si_3N_4 covered sample which had been ion implanted, was etched in HF to remove the Si_3N_4 layer. It was then annealed in arsenic at 800°C for 10 minutes. The metallization pealed on this sample during the patterning process. A probe of the surface showed ohmic behavior as is shown below with high resistance (~ 30,000Ω).



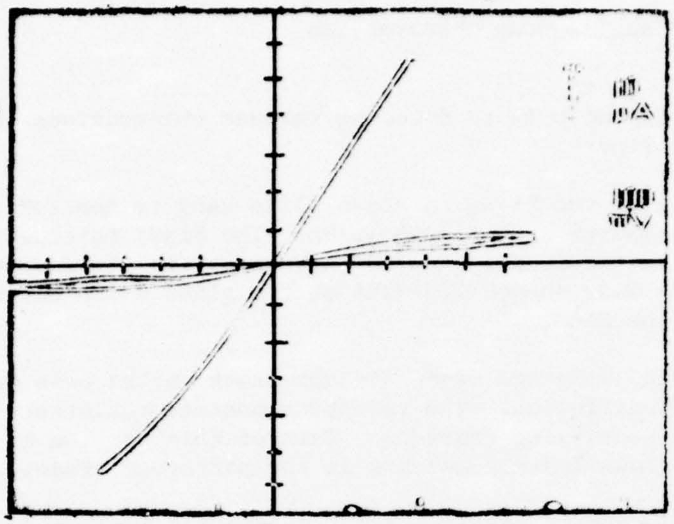
200μA/div vert

5V/div horiz

Sample C2X23.03

The sample was etched in HF after which it showed rectifying characteristics with no breakdown up to ~90 volts. After reprocessing the sample showed the I-V characteristics below. Further alloying of the contacts only made them more rectifying.

Microscope Light On



C23.03

Light Off

REFERENCES

1. Henish, H. K., Rectifying Semiconductor Contacts, Clarendon Press, Oxford (1957).

2. Willardson, R. K. and Allred, W. P., Gallium Arsenide, p. 35 (Edited by A. C. Strickland) Institute of Physics and the Physical Society, London, U.K. (1967).

3. Davies, D. E. et al, "The Role of Elevated Temperatures in the Implanation of GaAs," Solid-State Electronics 18, 733 (1975).

4. Kendal, D. K., Semiconductors and Semimetals, Vol. 4 (Edited by R. K. Willardson and A. C. Beer) Academic Press, New York (1968).

5. Young, A. B. Y., Stanford Technical Report No. 5116-1 (1969).

6. Asai, S. and Kodera, H., Proc. Third Conf. Solid State Devices, p. 231. Tokyo (1971).

7. Matino, H., "Reproducible Sulfur Diffusion into GaAs," Solid-State Electronics 17, 35 (1974).

8. Lane, T., Private Communication.

9. Cawley, A. M. and Sze, S. M., "Surface States and Barrier Heights of Metal-Semiconductor Systems," J. Appl. Phys. 36, 3212 (1965).

10. Pruniaux, B. R., "Transport Properties of the Gold Germanium Gallium Arsenide Metal Semiconductor System," J. Appl. Phys. 42, 3575 (1971).

11. Edwards, W. D., et al, "Specific Contact Resistance of Ohmic Contacts to Gallium Arsenide," Solid State Electronics 15, 387 (1972).

12. Yu, A. Y., et al, "Contacting Technology for Gallium Arsenide," Technical Report AFAL-TR-70-196 (1970).

13. Robinson, G. Y., "Metallurgical and Electrical Properties of Alloyed Ni/Au-Ge Films of N-Type GaAs," Solid State Electronics 18, 331 (1975).

14. Huang, C. I., "Contact Resistance of Planar TED," Memo Report dated 8 July 1975, AFAL/DHM-1.

15. Reeves, G. G. and Donovan, R. P., "Gallium Arsenide Technology, Volume II," Technical Report AFAL-TR-72-312, January, 1973.

16. Matino, Haruhiro, "Reproducible Sulfur Diffusion into GaAs," Solid State Electronics 17, 35 (1974).

17. Yeh, T. H., "Diffusion of Sulfur, Selenium, and Tellurium in Gallium Arsenide," J. Electrochem. Soc. 111, 253 (1964).

18. Personal Communications from J. Debesis, General Electric Solid State Lamp Project and from Osram Lamp Company, Munich, Germany.

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO

&

EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

A NAVIGATION ALGORITHM FOR THE LOW-COST GPS RECEIVER

Prepared by:	Philip S. Noe, Phd.
Academic Rank:	Assistant Professor
Department and University:	Department of Electrical Engineering Texas A&M University
Assignment:	
(Laboratory)	Avionics
(Division)	Reconnaissance & Weapon Delivery
(Branch)	Analysis
USAF Research Colleagues:	K. A. Myers and D. Botha
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

LIST OF FIGURES

A NAVIGATION ALGORITHM FOR THE LOW-COST GPS RECEIVER

by

Philip S. Noe

ABSTRACT

The global Positioning System (GPS) is a satellite navigation system currently under development by the Department of Defense. This note describes a new position fix algorithm for a low-cost GPS receiver. The algorithm uses Hotelling's method to iteratively update the inverse of the measurement matrix for correction of the navigation position. The last position estimate can be used to update the present position, so no dead reckoning procedure is required. Two modes of operation are considered: a stationary user/satellite system model and a moving user/satellite system model. Numerical convergence properties are tested for specified user position errors in the fixed model. Noise free and noisy models are considered in the dynamic system simulation. The basic convergence properties are tested in the fixed system, and convergence is assured with up to 6000NM initial displacement of the user. In the dynamic model without noise three iterations are required to obtain errors of the order of 10^{-6} ft. When noise is included in the model 400 ft. errors or less occur with convergence in 3 iterations. Similar results occur if the convergence is forced in one iteration.

INTRODUCTION

The Global Positioning System (GPS) is a satellite navigation system currently under development by the Department of Defense (Ref. 1). It will consist of 24 satellites in circular, 12-hour orbits at an altitude of 11,000 NM and inclined at 63° to the equator. The satellites will broadcast pseudo-random noise codes and ephemerides on two L-band signals to users worldwide. A user will be equipped with a small receiver (GPS user equipment) which measures pseudo-range and pseudo-range-rate from the user to the satellite. Typically, four satellite signals may be received simultaneously or sequentially by the user equipment. A major design consideration in the receiver is cost.

The purpose of this research is to present a navigation algorithm which could ultimately be implemented on a microprocessor chip as part of a low-cost GPS receiver. The algorithm requires a minimum of computational support from the user equipment hardware, but yet provides sufficient accuracy for a large number of potential GPS users; e. g., cargo aircraft and merchant marine vessels.

THE ALGORITHM

A fundamental problem in the implementation of a GPS low-cost receiver involves the computation of user position components u_1, u_2, u_3, and clock bias b which can be obtained from the spherical navigation equations:

$$\sum_{j=1}^{3} (x_{ij} - u_j)^2 = (r_i - b)^2 \qquad i = 1,\ldots,4 \tag{1}$$

where the x_{ij}, j=1,2,3 are the three known position components of the $i^{\underline{th}}$ satellite, and $r_i - b$ is one of the four pseudo-ranges measured simultaneously between the $i^{\underline{th}}$ satellite and the user. The conventional approach to this problem is to solve the four nonlinear equations in (1) simultaneously for the four unknowns in $u \equiv [u_1\ u_2\ u_3\ b]^T$; however, this direct solution is computationally unwieldy. It is the intent of this report to identify a simple approach which can be implemented in a microprocessor.

The approach here is to linearize (1) about the current estimate of user position and solve successively for position corrections based on new measurement residuals processed in the receiver. Rearranging (1) and solving for r_i gives

$$r_i = [(x_{i1}-u_1)^2 + (x_{i2}-u_2)^2 + (x_{i3}-u_3)^2]^{1/2} + b. \tag{2}$$

A Taylor series expansion of r_i is given by

$$r_i = \bar{r}_i + \frac{\partial r_i}{\partial u}\bigg|_{\bar{u}} \delta u + \frac{\partial^2 r_i}{\partial u^2}\bigg|_{\bar{u}} \delta^2 u + \ldots, \tag{3}$$

where $\bar{u}$ is the user's estimate of u and $\bar{r}_i$ is computed from $\bar{u}$ in (2). Now

(3) can be linearized bv droppinq all derivatives but the first, and the basic difference equation obtains as

$$\delta r_i = \left[\frac{\partial r_i}{\partial u_1} \quad \frac{\partial r_i}{\partial u_2} \quad \frac{\partial r_i}{\partial u_3} \quad \frac{\partial r_i}{\partial b}\right]\Bigg|_{\bar{u}} \delta u \tag{4}$$

where $\delta r_i \equiv r_i - \bar{r}_i$ and $\delta u \equiv u - \bar{u}$. Now define

$$\delta r_i \equiv h_i \ \delta u \tag{5}$$

where h_i is the row vector in (4). Thus, for the four-satellite case,

$$\delta r = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix} \delta u \equiv H \ \delta u \tag{6}$$

where H is the 4x4 matrix of partials of r with respect to u, evaluated on $\bar{u}$ and $\delta r \equiv [\ \delta r_1 \ \ \delta r_2 \ \ \delta r_3 \ \ \delta r_4]^T$. The desired position corrections are obtained by solving (6) as follows:

$$\delta u = H^{-1} \ \delta r. \tag{7}$$

Existence of the inverse is assured bv selection of four satellites with reasonably good geometry. The h_i in (6) are given by

$$h_i = \left[\frac{x_{i1}-u_1}{r_i-b} \quad \frac{x_{i2}-u_2}{r_i-b} \quad \frac{x_{i3}-u_3}{r_i-b} \quad 1\right]\Bigg|_{\bar{u}} \tag{8}$$

It is proposed that Hotelling's algorithm (Ref. 2) be used to iteratively calculate H^{-1} in (7) as the major tool in solving for the GPS position fix. Basically, Hotellinq's algorithm proceeds as follows:

$$G_m = G_{m-1}(2I-HG_{m-1}) \quad m = 1,2,... \tag{9}$$

where G_0 is some initial estimate of H^{-1}. The G_1, G_2, G_3,...successively approach H^{-1} if it can be shown that

$$||I-HG_0|| = ||F_0|| = k<1. \tag{10}$$

The number of correct digits in the inverse increases geometrically as

$$||G_m-H^{-1}|| \leq ||G_0||\frac{k^{2m}}{1-k}. \tag{11}$$

Here, the norm $||X|| \equiv \max_{i} \sum_{j=1}^{n} |X_{ij}|$. (12)

This algorithm provides a very simple solution to the GPS navigation problem. Use of the previous position estimate to compute H and G_0 in (9) and subsequent corrections in (7) to calculate G_1, etc. eliminates the need even for a dead reckoning position update between successive range measurement times. It is shown below that a significant amount of time can elapse for many user applications and still allow accurate computation of H^{-1} for insertion in (7). Position error constraints and hence time and velocity constraints are discussed such that (10) can still be satisfied.

The algorithm is implemented in two forms one of which allows G_m to converge to maximum accuracy before updating $\bar{u}$ with the change in δu computed in (7), and a second form in which only one iteration of G is allowed before updating $\bar{u}$. Convergence is controlled and determined by comparing δu in (7) with an arbitrary minimum value and $G_m - G_{m-1}$ with this same arbitrary minimum value. Hence the algorithm iterates in two ways:

$$G_k = G_{k-1}(2I - HG_{k-1}) \tag{13}$$

and

$$\bar{u}_k = \bar{u}_{k-1} + \delta u_k. \tag{14}$$

If $|G_k - G_{k-1}| < \varepsilon$ and $|\delta u_k| < \varepsilon$, then the algorithm has converged (here ε is an arbitrary minimum value). It is shown below that it is unnecessary to let G_k converge to maximum accuracy in (13) at each iteration of (14) since G_k is not in its final form until (14) has converged anyway. In summary the algorithm steps are stated below:

Algorithm

```
1) Guess initial $\bar{u}$; select 4 satellites; compute $G = H^{-1}$
2) Obtain $r_i$ and $x_{ij}$ from receiver data
3) Calculate $\bar{r}_i$ for all i from (2)
4) Obtain $\delta r = r - \bar{r}$
5) Obtain the H matrix from (8)
6) Obtain $G_m$ from (9)
        If $|G_m - G_{m-1}| > \varepsilon$ repeat step 6); otherwise, go to 7)
7) Obtain $\delta u = G \delta r$
8) Calculate $\bar{u}_k = u_{k-1} + \delta u_k$
        Repeat steps 2) through 8) if $\delta u_k > \varepsilon$ ; otherwise, stop.
```

SIMULATION MODEL

Two models are used for simulation: a fixed model and a dynamic model. The fixed model consists of fixed user and fixed satellites. The primary purpose of the fixed model is to determine the maximum user displacement under which convergence of the algorithm will occur. The purpose of the dynamic model is to illustrate convergence properties of the algorithm in a realistic flight simulation of a C5A aircraft.

Fixed Model Simulation

In the discussion above, it is stated that convergence will be investigated with respect to time differences between successive fixes. A more general parameter affecting convergence is the displacement distance d. If the maximum distance d_{max} for convergence is known then the set of all possible velocity-time variations can be directly determined on a hyperbolic curve given by $d_{max} = vt$, although other factors may affect convergence in a dynamic model.

It is clear that d_{max} exists as a result of the constraint indicated in (10). It is not clear in the statement of the geometric convergence properties of (11) how $k<1$ affects the number of iterations before convergence occurs. One objective of this paper is to determine the number of iterations required for convergence and the corresponding computation time on a CDC 6600. These results can then be used to obtain an indirect estimate of the feasibility of a microprocessor implementation and the resultant velocity constraints for a particular class of users.

The simple stationary user/satellite model shown in Figure 1 is used to numerically obtain an estimate of d_{max} where divergence of the algorithm occurs. The user is fixed on a spherical earth at $x_1=x_2=0$ and $x_3=3440$ NM (earth fixed rectangular coordinates) with four stationary satellites, one of which is "overhead" and the other three are 120° apart on the user's horizon as indicated in the figure. Reasonably good satellite geometry has been assumed so that user position errors δu are observable. It is felt that this model is adequate to validate the general numerical convergence properties of the algorithm. The user's initial position estimate $\bar{u}$ is generated by perturbing the true position, u, by some specified parameter, d; i. e.,

$$\bar{u} = u + d. \tag{15}$$

The perturbation d is increased linearly until the algorithm fails to converge to the true position u, and the last converging value of d corresponds to d_{max}.

Dynamic Model Simulation

Satellite position and C5A mission profile data are generated in a realistic computer simulation program (SATGEN and PROFGEN respectively) on the CDC 6600 computer. SATGEN simulates the propagation of the 24 GPS transmitters (satellites) in circular orbits about a point mass earth model. Two modes of operation are employed in SATGEN: 1) all satellites are tested for visibility with respect to the user's estimated position and subsequently a suboptimal procedure is used to select four satellites which provide good observability of user position error; 2) position data for these four satellites is updated as a function of time. The algorithm used to select the suboptimal set of satellites consists of two parts: 1) select the satellites with maximum range in each of the three coordinate directions, and 2) from the satellites remaining select the fourth satellite which has the best GDOP (Ref. 3). Although this algorithm is suboptimal the satellite set selected frequently is optimal and has never been "bad". Computational efficiency is the major advantage of this pseudo-optimal procedure. If n satellites are in view, $\binom{n}{4}$ GDOP computations are required to determine the optimal set of satellites. On the other hand the suboptimal method requires only $n-3$ GDOP computations. In general this is quite a computational advantage. A detailed statistical analysis of the satellite selection algorithm is beyond the scope of this research.

Frequency of satellite selection is clearly a function of time and user's velocity in order for the satellites to remain in view and to insure that the GDOP is good. For the purpose of this simulation it was experimentally determined that new satellite selection is required every 15 minutes. The satellite generation routine was tested with a fixed user simulation where new satellites were selected every three minutes. The displacement d was fixed at 100 NM, and an hour of fixes were taken. Although totally new satellite sets were often taken with large variations in user to satellite ranges, re-inversion of the H matrix was not required. Typically the selected set of satellites remained the same for well over 15 minutes; thus, the 15 minute selection time was established experimentally.

The incorporation of a moving user at a 600 MPH velocity in conjunction with a 15 minute satellite selection time showed that it is necessary to reinvert the H matrix at the time of satellite selection; otherwise, computation of G at Step 6) diverges whether internal iteration at Step 6) is allowed or not. Hence, in this dynamic model the H matrix is reinverted at the time of satellite selection.

The satellite position data generator and mission profile data generation routine PROFGEN were developed independently by the USAF sponsors. The PROFGEN routine is capable of generating mission profile data for presumably any aircraft dynamics and produces data representing a simulated flight from any point A to point B. In addition to the above routines the USAF sponsor also provided the complete noise model for the noisy mission data simulation.

FIXED MODEL SIMULATION RESULTS

Results for the fixed model are surprising. A plot of displacement versus the number of iterations for convergence is shown in Figure 2 for $d \leq 6050$ NM. The value of $d_{max} = 6050$ is considerably higher than expected. Convergence is initially controlled by two factors: internal iteration in the Hotelling equation at Step 6) and subsequently convergence of $\bar{u}$ in (14) at Step 8) such that $|\delta u| \leq 10^{-8}$. Thus, a two step iteration procedure is recursively implemented which produces convergence in 12 to 36 iterations. This cutoff criterion leads to position component errors of the order of 10^{-10}NM. It should be noted that as long as the satellite configuration has what is known as a "good" GDOP (geometric dilution of precision--Ref. 3), there is little change in d_{max}. Many other satellite configurations (effectively a moving satellite model with fixed user) have been investigated with comparable results.

Next, the basic algorithm is considered without iterating (13) at step (6) in an effort to reduce the total number of iterations required for convergence. It was found that fewer total iterations are required if iterations internal to the Hotelling equation are precluded. This procedure is valid since the user's position estimate is only approximate at Step 8) until terminal convergence occurs anyway. This modification produces convergence in 3 to 5 iterations for 100 different fixes with $10 < |d| \leq 1000$NM. Subsequently, a more realistic cutoff criterion of $|\delta u| \leq 10^{-5}$ NM was used to see if a small position error remains with fewer iterations. For 1600 fixes with $|d| \leq 1000$NM only 3 or 4 iterations were required, position error is $\leq 10^{-10}$ NM, and d_{max} is still of the order of 6000NM as shown in Figure 3. The impressive feature of the modified algorithm is that only 3 or 4 iterations are required for initial position errors of $|d|\leq 1000$ NM. A typical computer run on the CDC 6600 requires from 20 to 75 msec/fix.

DYNAMIC MODEL SIMULATION RESULTS

The C5A flight is simulated with and without noise in the data generated by the satellite and mission profile routines. Fixes are obtained every 30 sec., and the satellites are selected every 15 min. For the noise free case a total of 720 fixes are taken during the 6 hour flight with a typical error of 10^{-6}ft. or less. The algorithm requires 3 iterations to converge for each fix, and the convergence criterion for iteration is $\delta u \leq 10^{-5}$ ft. in Step 9) of the algorithm.

An additional noiseless flight was made with $\delta u \leq 10^{5}$ as the convergence criterion. With this value for δu only one iteration is required for each fix, and the errors are 8 ft. or less. A plot of error versus time is shown in Figure 4 for this single iteration case. Similar results are obtained if the optimal criterion for GDOP is used for satellite selection. Errors of less than 8 ft. are obtained, and these results are shown in Figure 5.

For the noisy data case convergence is obtained in 3 iterations although the errors are 300 ft. or less as shown in Figure 6. A similar run with one iteration produces errors of 340 ft. or less as shown in Figure 7. Results are also shown for these two cases with the optimal GDOP criterion used for satellite selection. These results are shown in Figures 8 and 9. It is clear from these figures that there is no extensive variation in these results and those obtained by the suboptimal selection procedure.

It was experimentally determined that the convergence of the algorithm is a direct function of satellite selection time and fix time interval as discussed briefly in the dynamic simulation model. If H is not inverted at satellite selection time in the 15 min. select/30 sec. fix case, the algorithm does not converge. On the other hand, if a 30 sec. select/5 sec. fix case is considered convergence does occur without reinverting the H matrix at satellite selection time. Further research is needed to determine the exact bearing of these two parameters on convergence.

It is clear that the navigation problem can be solved without using the Hotelling algorithm (Step 6). As an alternative, G, the inverse of H, can be calculated by direct inversion. Two separate computer runs were made with 10,000 operations of the two methods. The Hotelling algorithm shows a 34% time advantage over direct inversion. This test clearly demonstrates the efficiency of the algorithm of this report.

CONCLUSIONS

It has been shown that the proposed navigation algorithm is a fast, accurate, convergent routine that offers sufficient position determination accuracy for a variety of low-cost GPS users. Therfore, it should be considered for implementation on a microprocessor chip as part of a low-cost GPS receiver. Application of this algorithm to derive position corrections for the more expensive high dynamic GPS receivers and for inertial system aiding is a topic for futher research.

References

1. Harold Shoemaker, "The NAVSTAR/Global Positioning System," Proceedings of the NAECON 75 conference, Dayton, Ohio, 10-12 June 1975.

2. V. N. Fadeeva, Computational Methods of Linear Algebra, Translated by C. D. Benster, Dover Publications, Inc., 1959, New York.

3. H. B. Lee, "A novel procedure for assessing accuracy of hyperbolic multilateration systems," IEEE Trans. Aerospace and Electronic Systems, vol. AES-11, pp 2-15, January 1975.

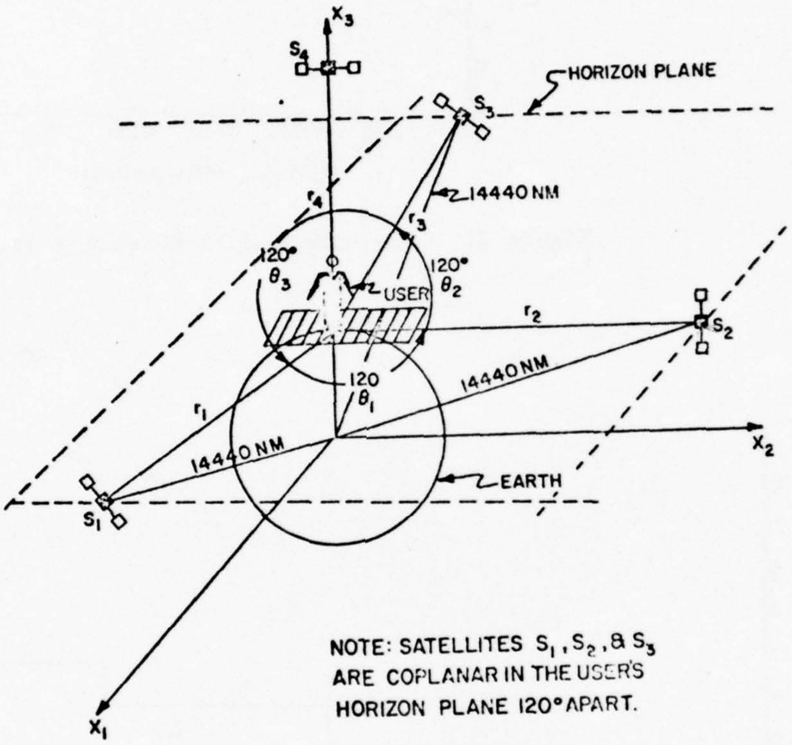


Figure 1. A Simplistic Satellite/User Model

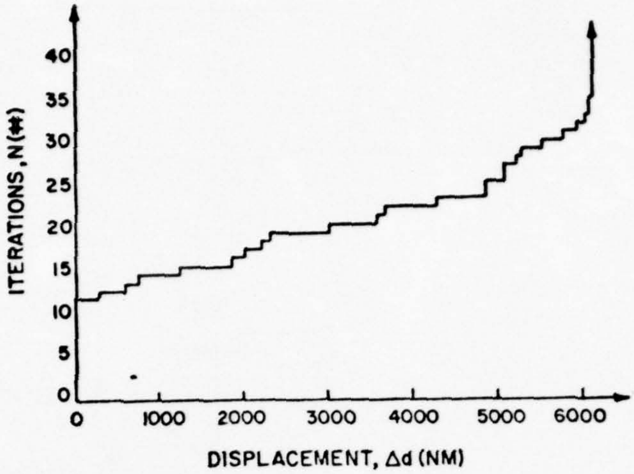


Figure 2. Convergence Data-Hotelling Iterating and δu

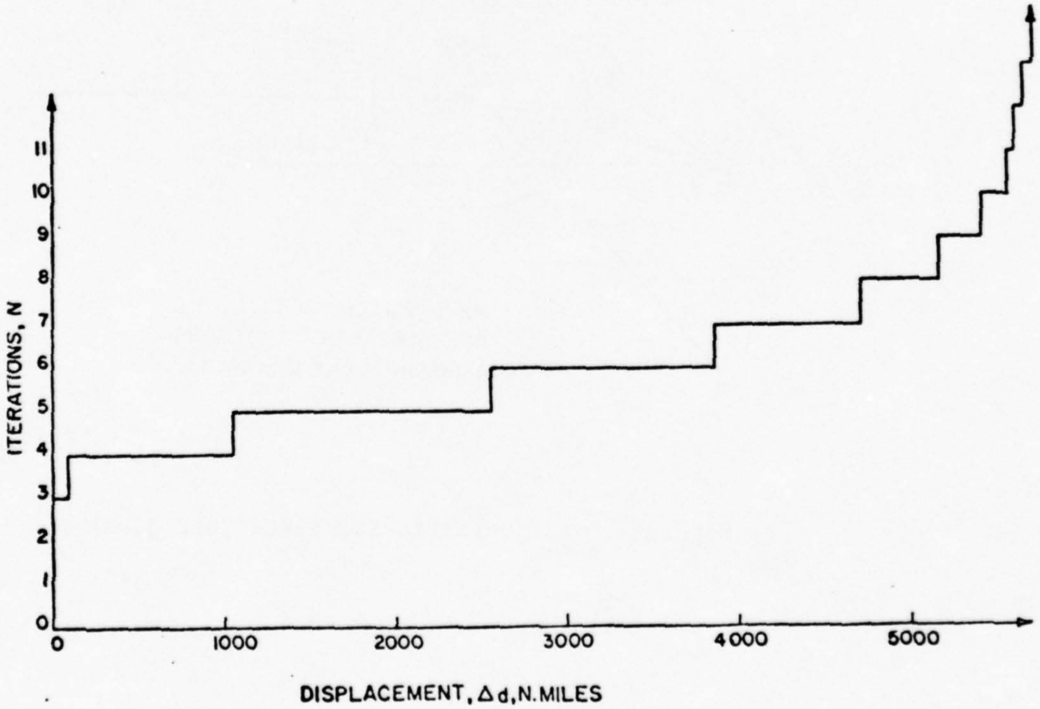


Figure 3. Convergence Data-δu Iterating Only

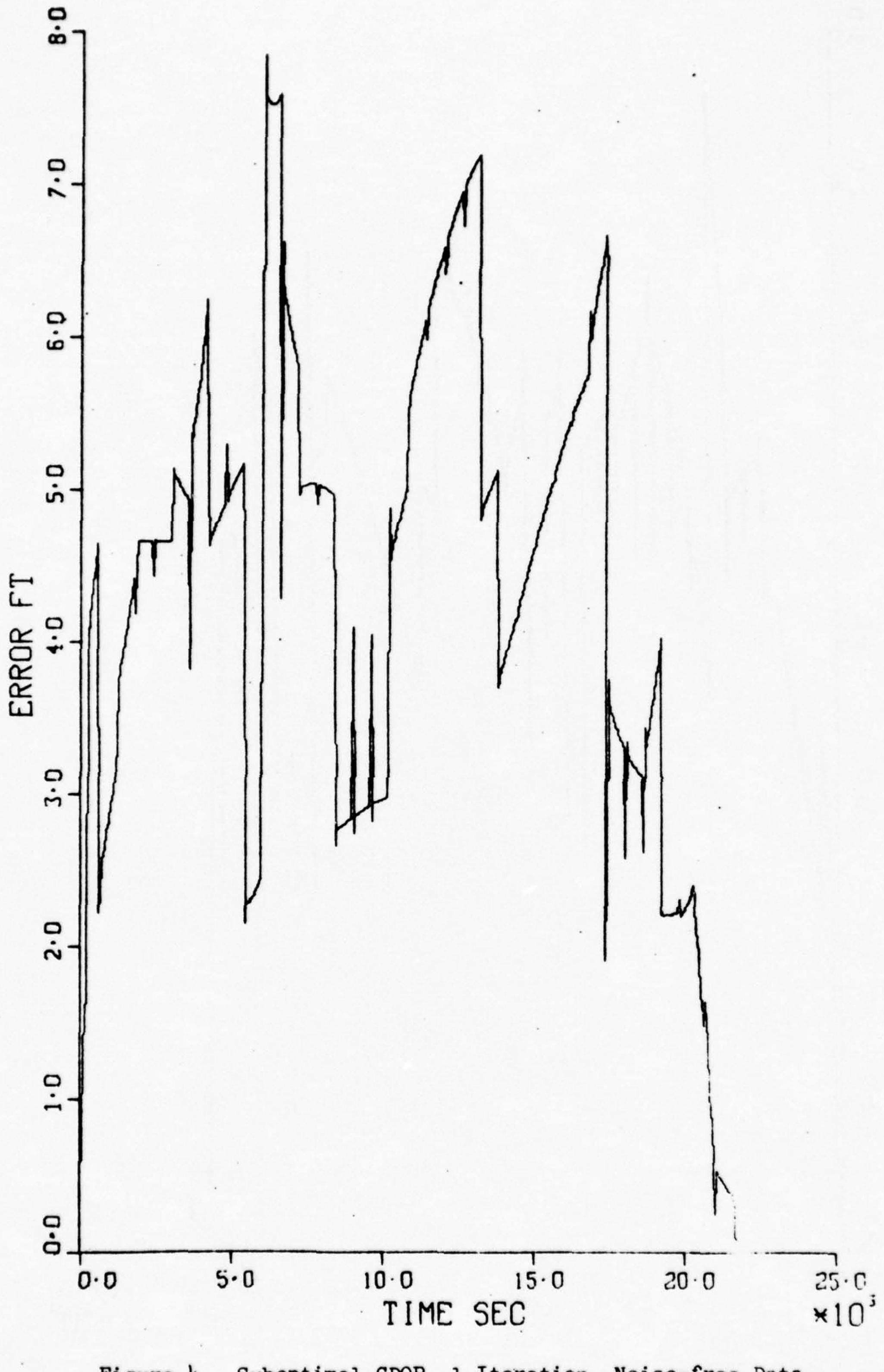


Figure 4. Suboptimal GDOP, 1 Iteration, Noise-free Data

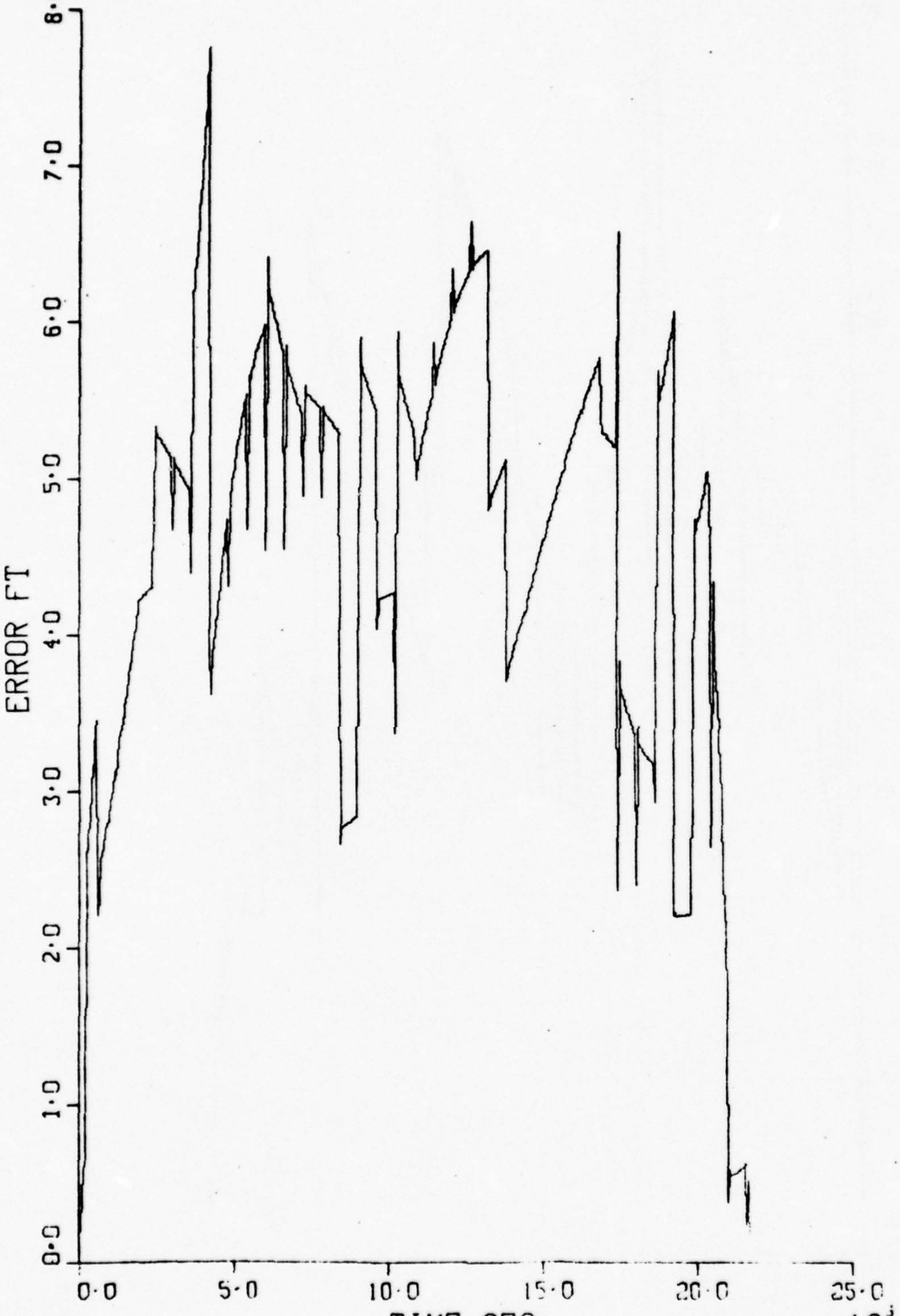


Figure 5. Optimal GDOP, 1 Iteration, Noise-free Data

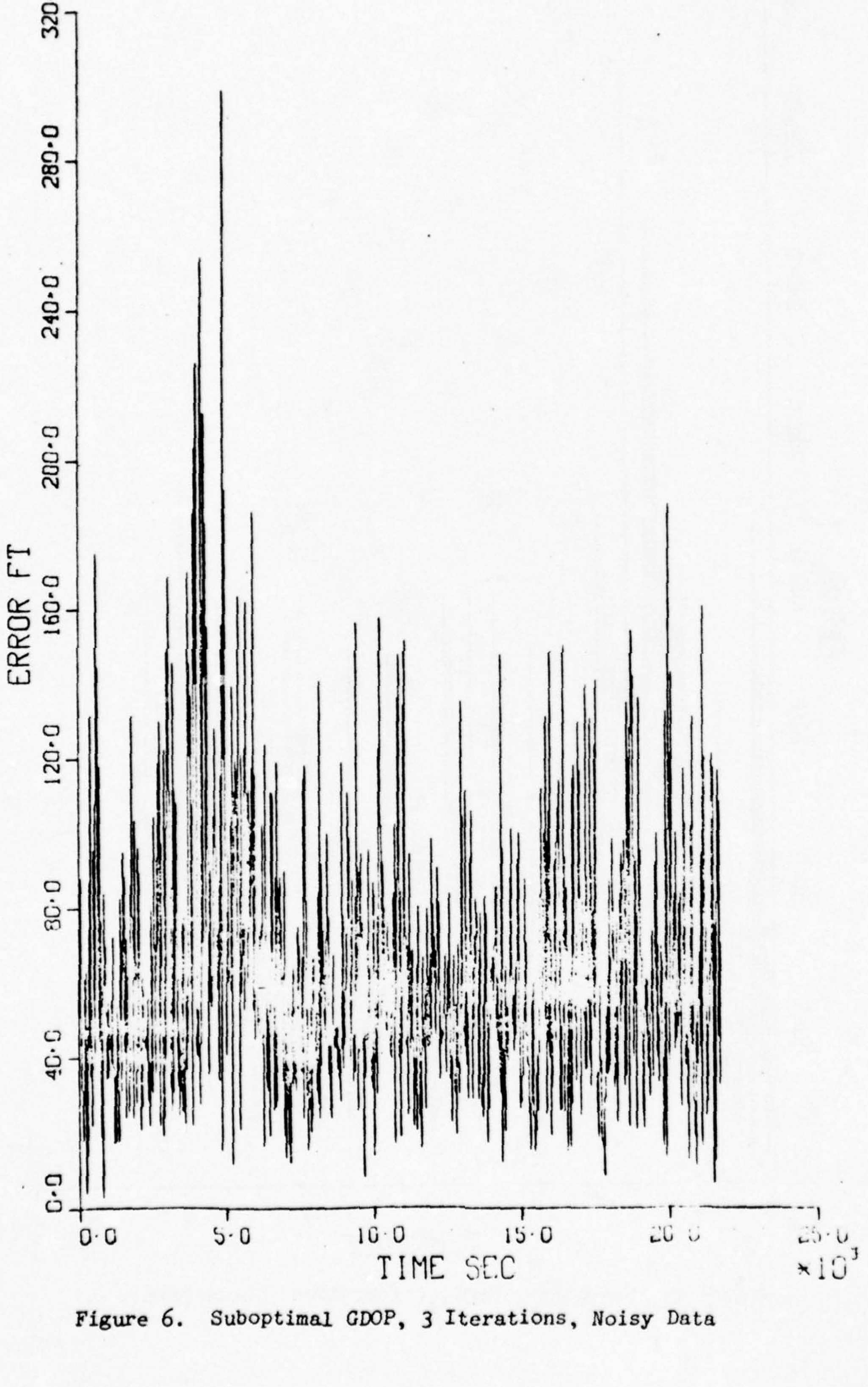


Figure 6. Suboptimal GDOP, 3 Iterations, Noisy Data

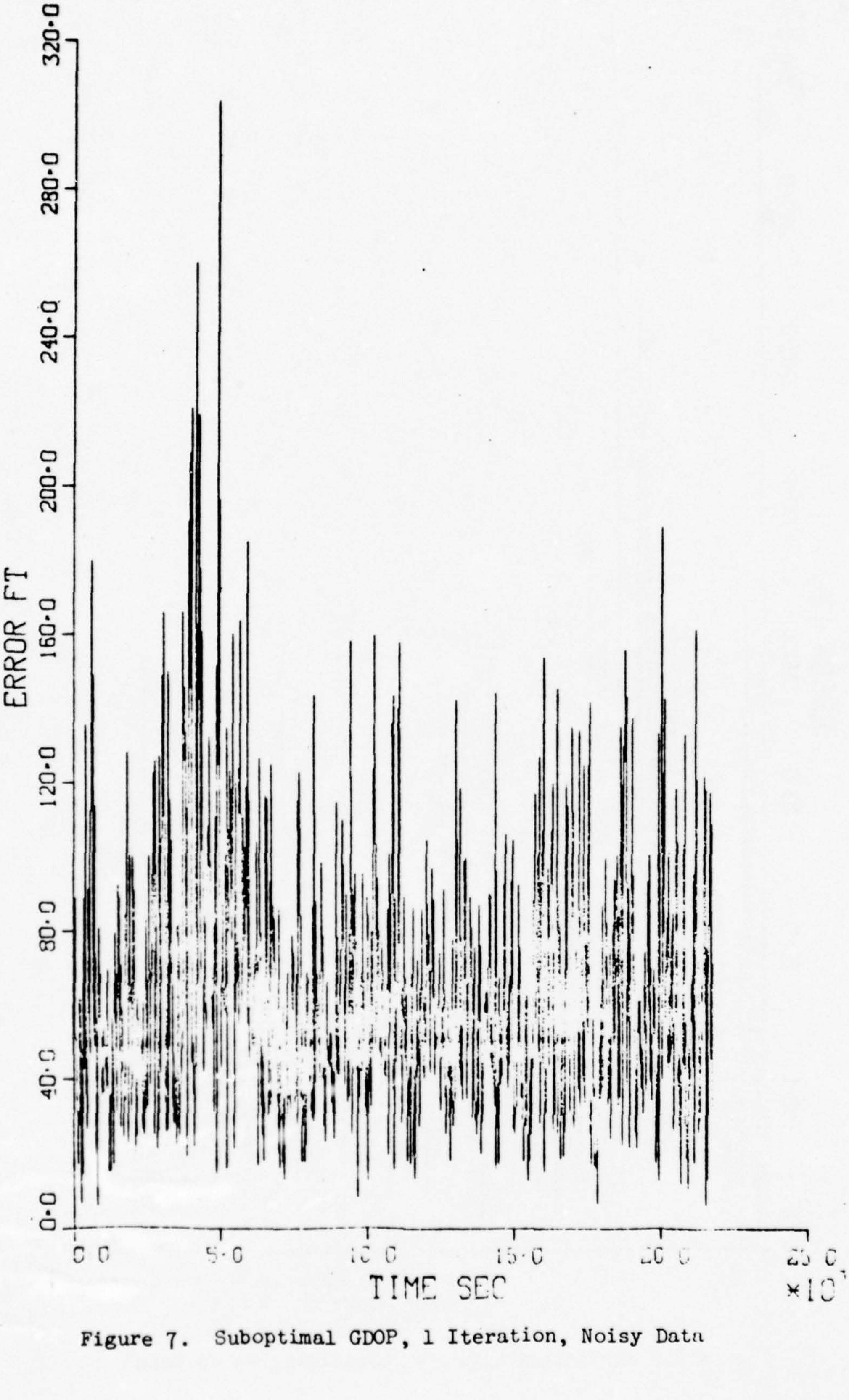


Figure 7. Suboptimal GDOP, 1 Iteration, Noisy Data

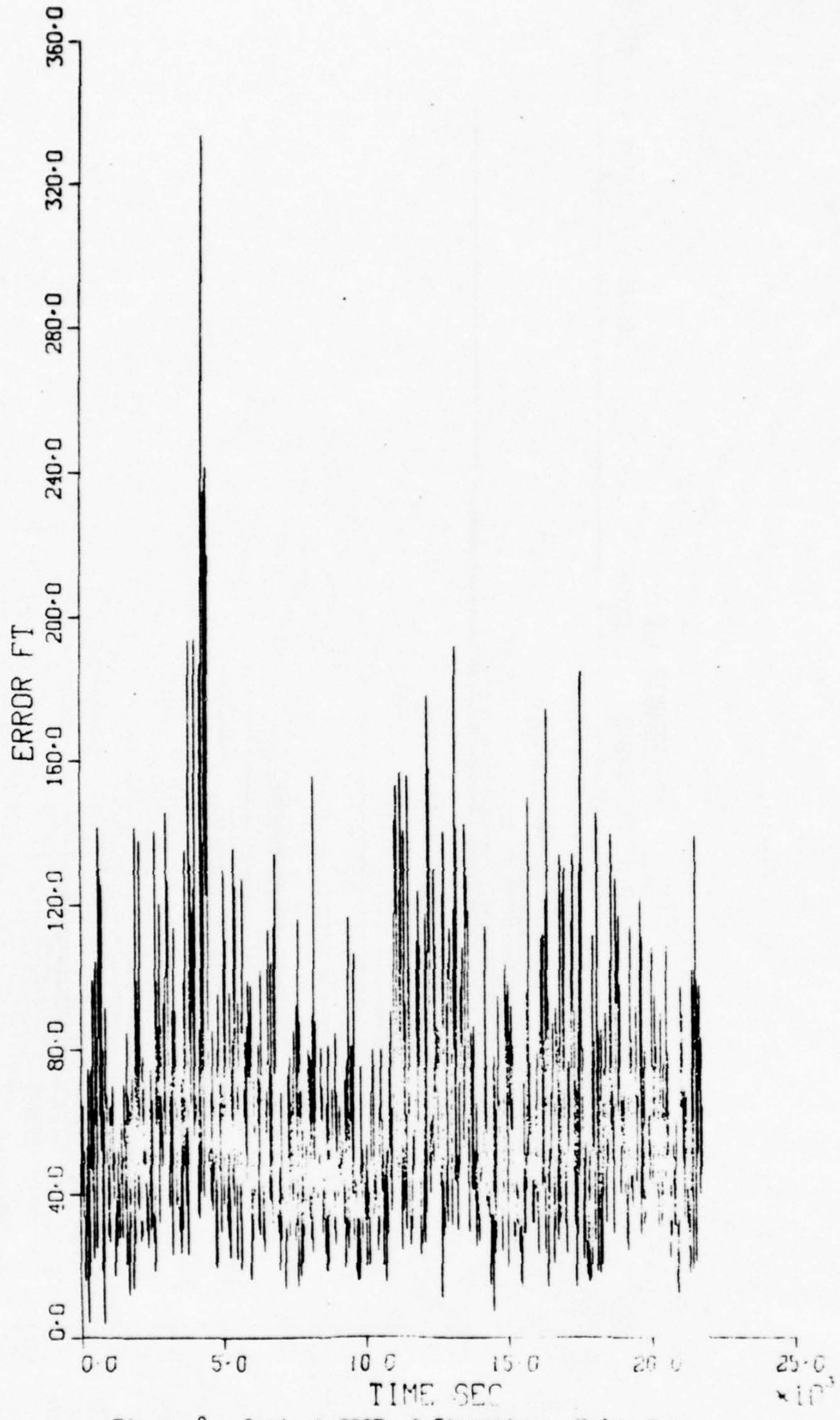


Figure 8. Optimal GDOP, 3 Iterations, Noisy Data

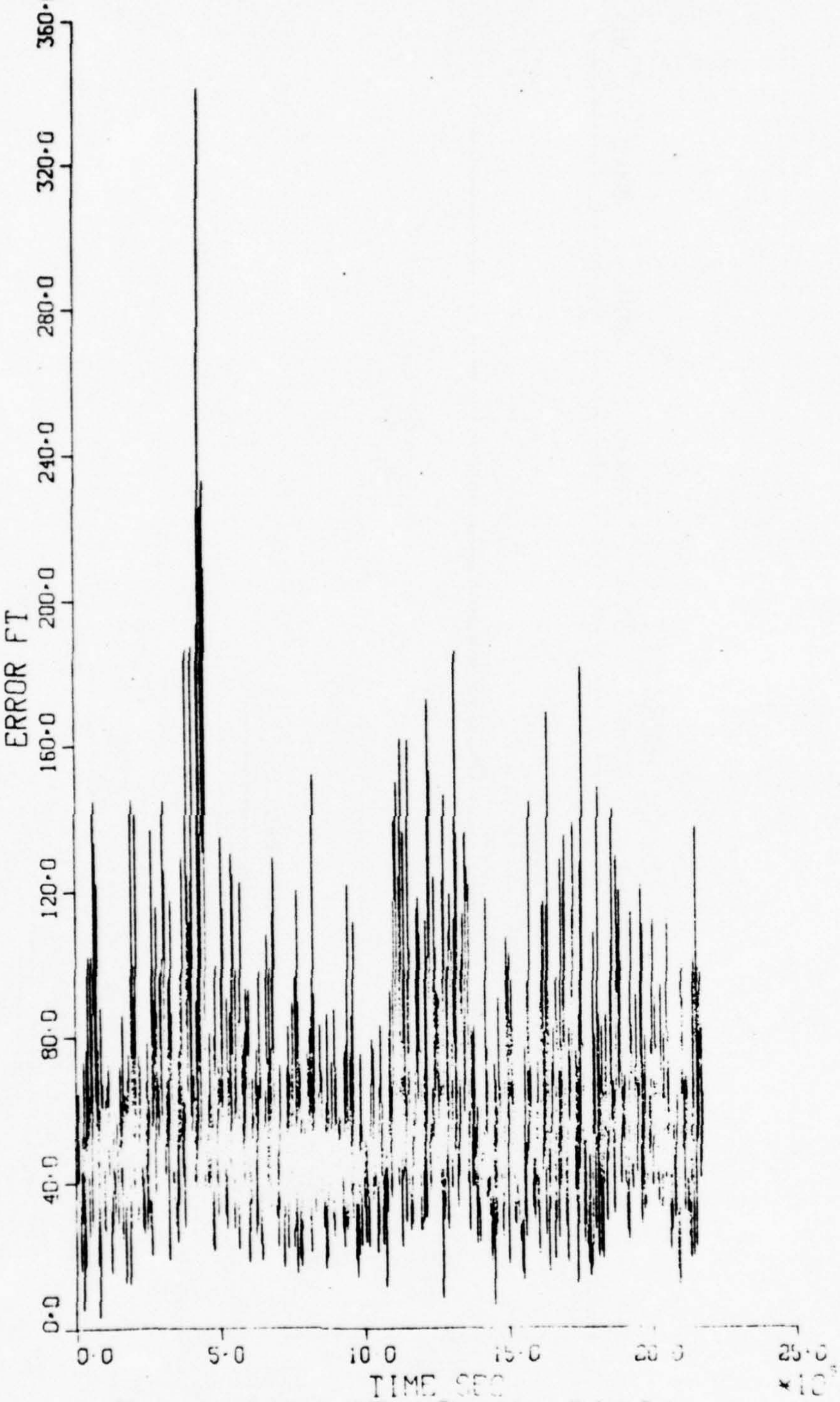


Figure 9. Optimal GDOP, 1 Iteration, Noisy Data

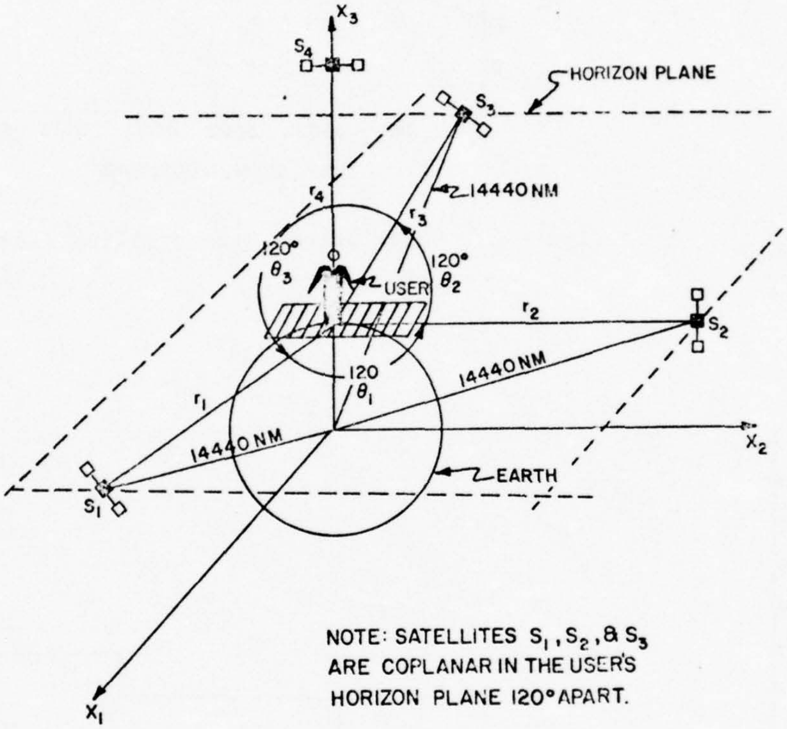


Figure 1. A Simplistic Satellite/User Model

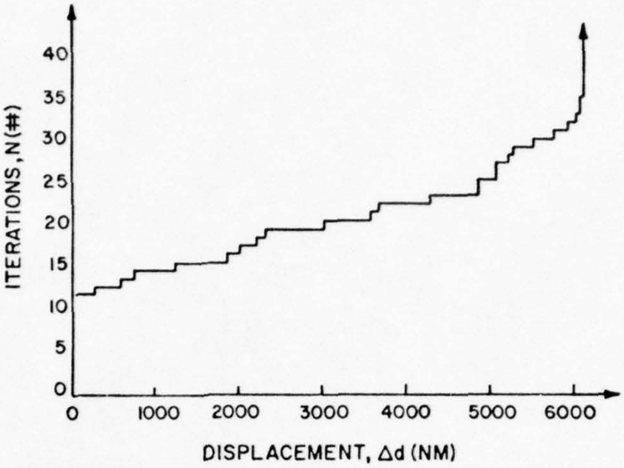


Figure 2. Convergence Data-Hotelling Iterating and δu

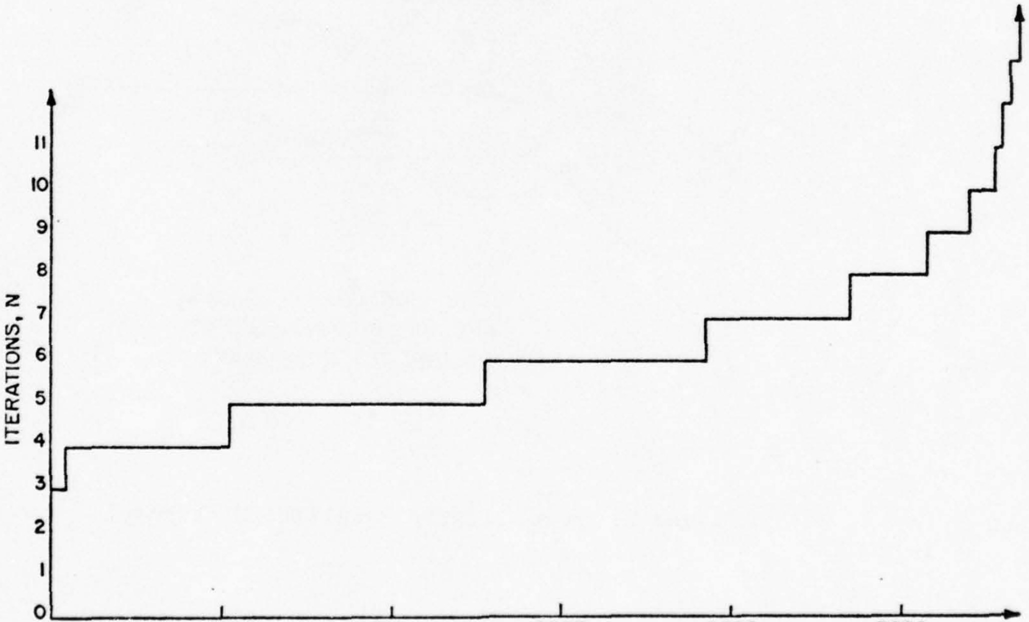


Figure 3. Convergence Data-δu Iterating Only

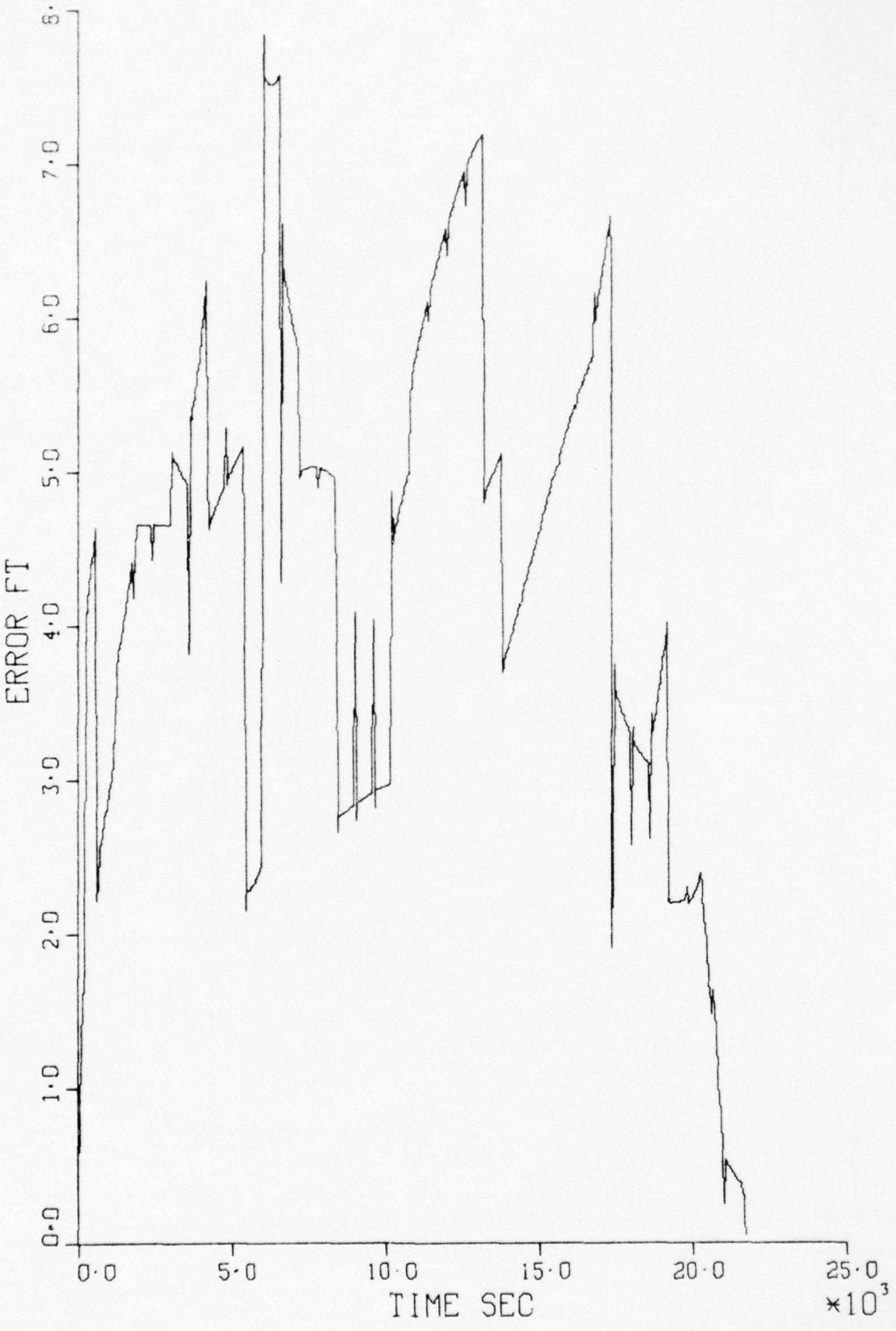


Figure 4. Suboptimal GDOP, 1 Iteration, Noise-free Data

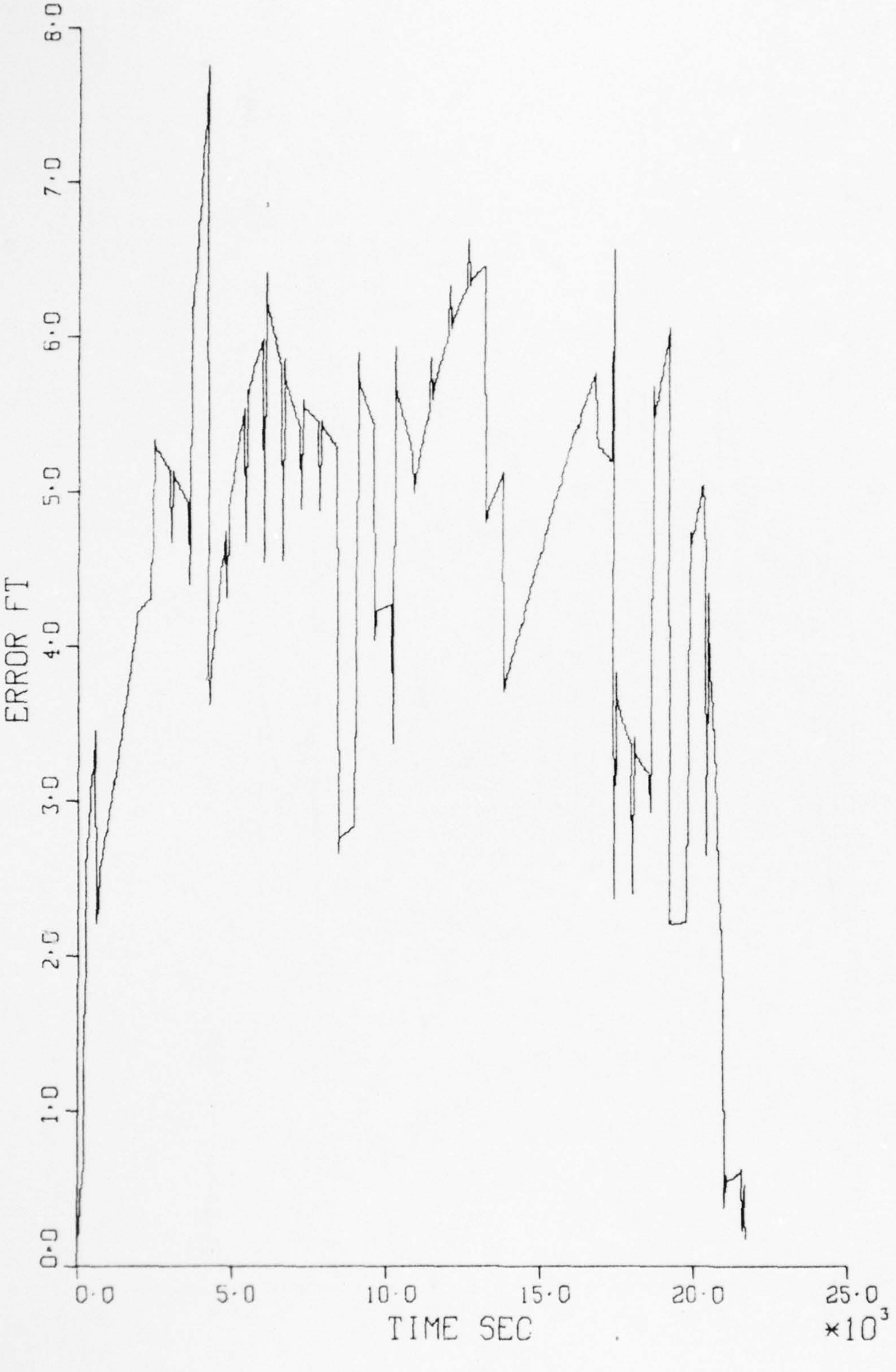


Figure 5. Optimal GDOP, 1 Iteration, Noise-free Data

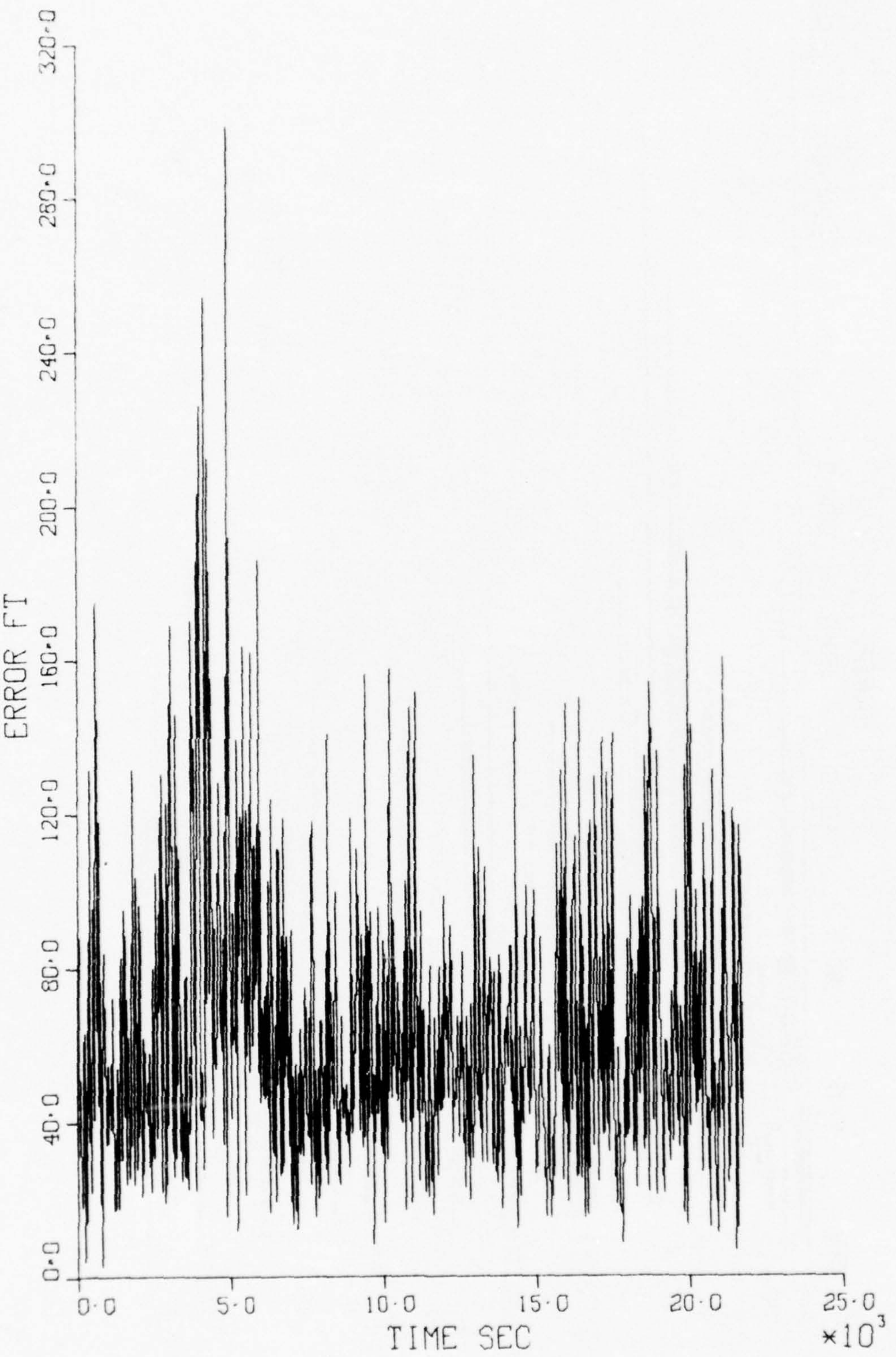


Figure 6. Suboptimal GDOP, 3 Iterations, Noisy Data

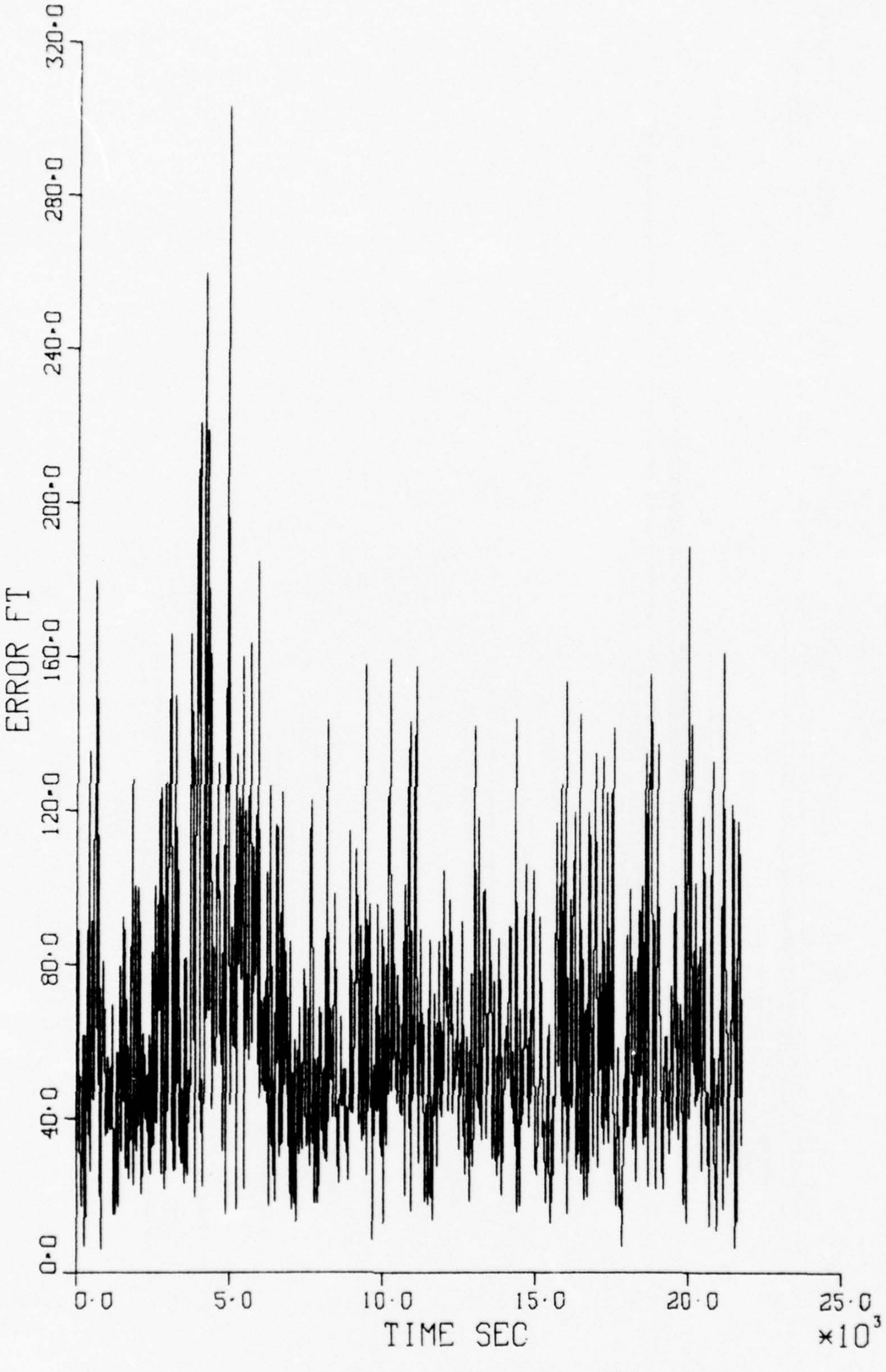


Figure 7. Suboptimal GDOP, 1 Iteration, Noisy Data

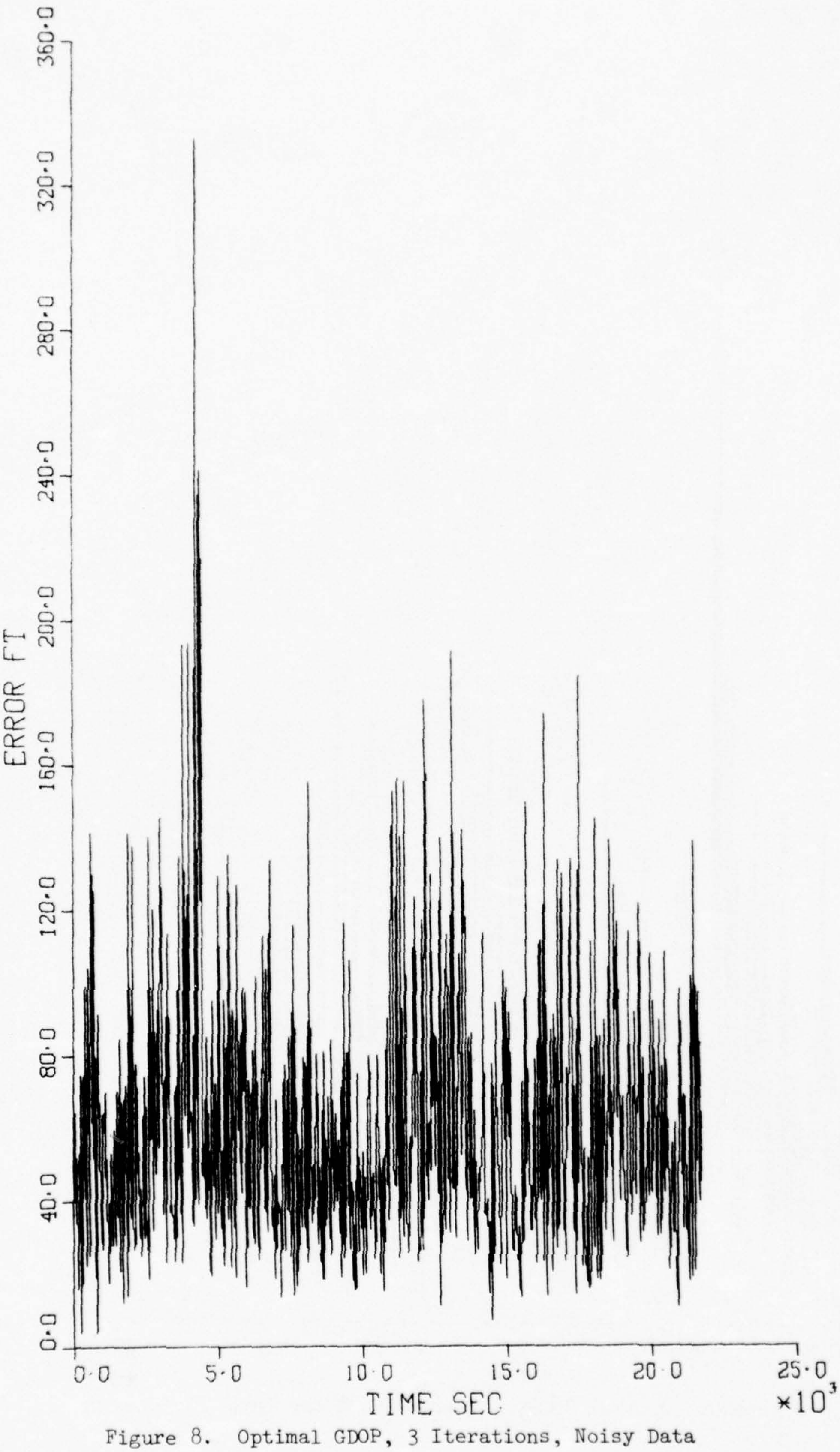


Figure 8. Optimal GDOP, 3 Iterations, Noisy Data

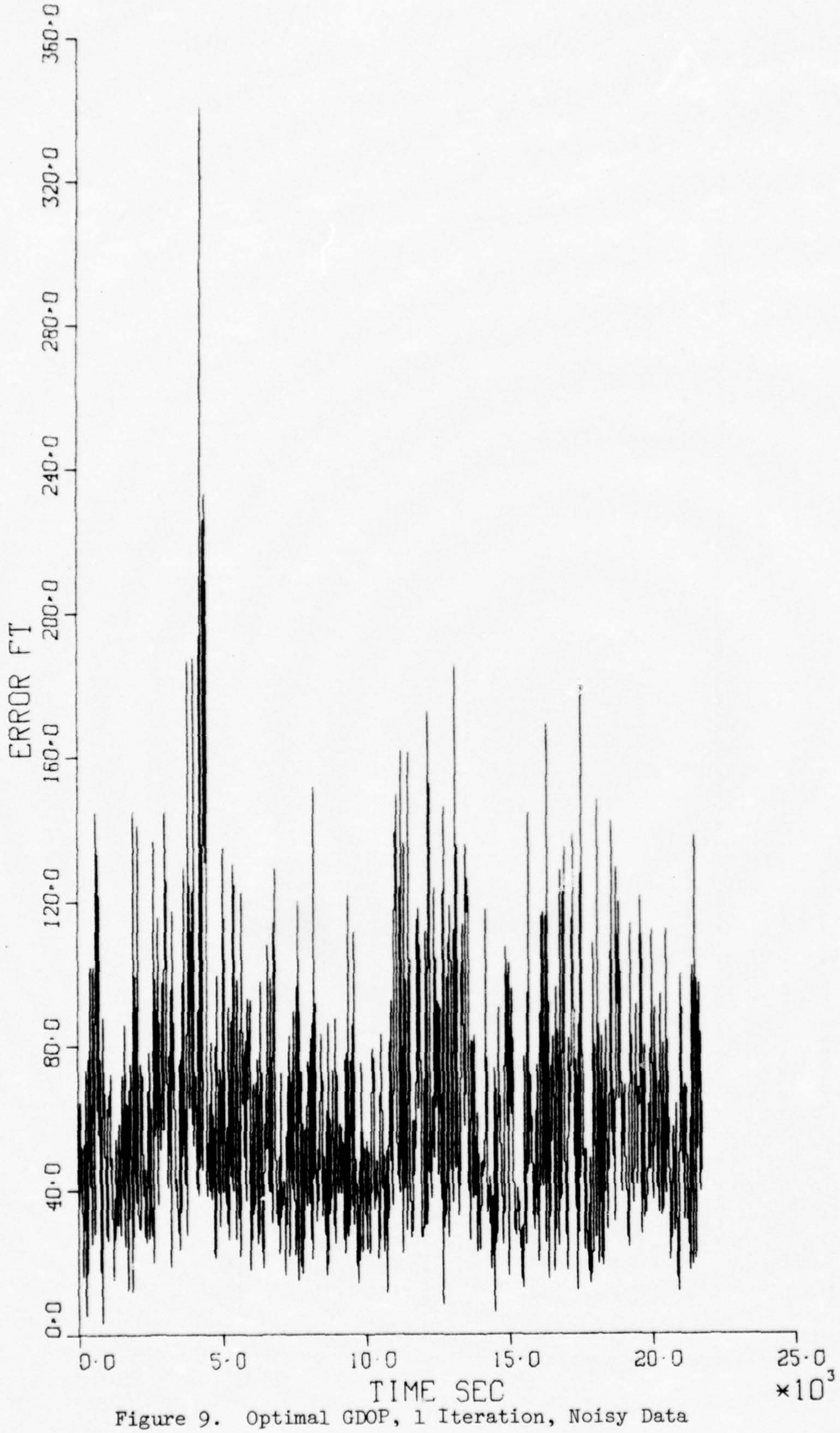


Figure 9. Optimal GDOP, 1 Iteration, Noisy Data

1975

ASSE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

ELECTRO-OPTICAL TRACKER ANALYSIS

Prepared by:	Jerry W. Rogers, PhD.
Academic Rank:	Associate Professor
Department and University:	Department of Electrical Engineering Mississippi State University
Assignment:	
Laboratory	Avionics
Division	Reconnaissance and Weapon Delivery
Branch	Analysis and Evaluation
USAF Research Colleague:	Capt Gary Reid
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

ELECTRO-OPTICAL TRACKER ANALYSIS

By

Jerry W. Rogers

ABSTRACT

A Honeywell electro optical tracker which uses a RETICON (50x50 diode array) was tested on the zoom optical target simulator (ZOTS) located at Holloman Air Force Base. The data acquired consisted of nine nine-track magnetic tapes. Results of data reduction are supplied.

Plots of position error were made as functions of time. Several of these plots were made where irradiances and contrasts were varied. A print out of selected digitized video is also furnished.

INTRODUCTION

Solid state imaging devices such as the diode array, charge coupled (CCD's), and the charge injection (CID's) have properties which are, for some applications, superior to the standard vidicon. For example, these devices are small, rugged, expected to be inexpensive and self scanning. When taking these properties into consideration, it becomes evident that potential benefits exist in optical tracking systems, particularly in air borne operations. This potential served as the impetus for testing an optical tracker with a solid state imaging device as the primary sensor. Because there are three devices all of which are different, it becomes necessary to investigate the relative merit of each. The first test in a series of tests is the integration of a diode array (RETICON) into an optical tracker built by Honeywell. The data described in this report comes from this test. Subsequent tests involving the CCD's and CID's have been conducted, but it will not be included in this report.

The tests were conducted at Holloman Air Force Base on their zoom optical target simulator (ZOTS). A description of this facility is given in Appendix A.

DATA FORMAT AND EXTRACTION

The data tapes were recorded on nine reels, each of which was a nine-track tape with 800 bits per inch in IBM compatible, NRZI format. Each frame consists of 24 computer words of 16 bits each. Each record contains 40 frames. The frame rate for most tests were 92.6 fr/sec and 23.1 fr/sec although some of the final runs were made at 185.2 fr/sec.

In order to make the data palatable to the CDC 6600, some bit shifting from the HP 16 bit word to the CDC 6600 60-bit word was necessary. The procedure employed was to select particular runs based on summary report, convert the records to a convenient CDC 6600 format and use the CDC 6600 permanent file capability for temporary storage (7 days). This obviated the exercise of physically mounting the nine track tape every time; a time consuming operation. Other programs were written which extracted information and displayed it in a digestable form.

SYSTEM MEASUREMENTS SELECTED

Because of the amount of data furnished and the limited time, it was necessary to pick and choose meaningful data for extraction. In fact, not all meaningful data was extracted; only that which could be considered in limited time was finally chosen.

The target on the simulator, which the tracker was to follow, was a opaque square. The square covered a 3 X 3 diode matrix. The motion was constrainted to 3 directions. One test was entirely along the X axis, one entirely along the Y axis, and one diagonally. The motion was oscillatory, constant velocity which would result from a triangular excitation of the servo drives.

The target motion along X was arbitrarily selected for observation. A look at system performance was made as irradiance and contrast was varied. During these tests, target velocity and frame rate were fixed. Plots of the error function and the digitized video print out were obtained.

RESULTS OF DATA REDUCTION

As a first step, all 107 records of file 1 were processed by the CDC 6600, and a print out of manual data, record number, frame number, time count, tracker X integer, tracker Y integer, ZOTS X integer, ZOTS Y integer, and the corresponding X physical positions in inches was obtained. This is a rather volumnous print out (100 computer pages) and will not be included in the report. Observation of these data show the servo to be noisy in and about zero displacement. The "Y" axis has a maximum noise of .0028 inches (about 70% of a physical cell distance). In some places it is periodic at 5 Hz and 30 Hz in other places. Because the motion is less than a cell distance, the tracker position never changed; however, a more dense array would produce a "chatter."

Figures 1 through 5 show the time history of the position error which is defined to be ZOTS' position (actual) - tracker's position on the X axis. This calculation was made at each frame and extended through twenty four (24) records; each curve is the result of 960 points. For these curves, the irradiance varied from a maximum of 21.0 μwatts/cm^2 to a minimum of 1.74 μwatts/cm^2 while the contrast remained high (Cont. > 56%). The errors vary from a maximum of 1 cell to minimum of zero. These maximum variations occur generally at the tracker decision-to-move frame. The greatest error occurs at the array edge where the target changes direction. This, of course, is expected in that a correlation tracker must update. There is one worrisome detail of this curve in that the "eye ball" average seems greatest at maximum target excursions. An additional plot, Figure 6, of error versus target position confirm the observation. Ideally, the curves in Figure 6 should have been rectangular. It is uncertain at this point whether this is a tracker phenomena or faculty calibration of ZOTS equipment. Figure 5 which was taken at low irradiance (1.74 μwatt/cm^2) shows a loss of lock around 9 seconds. One concludes that tracking is marginal at levels of 5 to 6 μwatts/cm^2 of lower. Figures 3 and 4 do not on the surface seem to confirm this conclusion, but notice that these two curves are made with high contrast. Figure 7 through 10 furnish more information along these lines.

Figures 7 through 10 show the results of contrast variation. They indicate that the peak error from 39% down is in excess of one cell. A digital print out of video shown in Figures 11 through 13 are also of interest in this respect. The print out shows the 5 X 5 array which the tracker uses for encompassing the 3 X 3 used for correlation. Notice that the 3 X 3 center elements are of lesser value than the outer 16 elements. Several frames are included vertical in order to show the tracker shifting positions. Presumably, these digits are in some sense related to actual video amplitude. Qualitatively, the differential from center to outer edge is constant. It is seen that by decreasing contrast by 50% at higher contrast levels show greater numerical change in the integers than doing a similar change at lower contrast levels. It appears that the video magnitude differential decreases with decreasing contrast, but not linearly. This condition does not favor low contrast targets.

In summary, several unresolved problems need further attention. They are not for the most part very serious problems, but are more of the nature for clearing up details which relate to the quality of test results. They are as follows:

1. Determine the extent of X-axis oscillations with zero input
2. Resolve the X-axis error anomaly
3. Determine the reason for the non-linearity in contrast changes
4. Look more specifically at the update rates for low contrasts
5. Look into the possibility of including zoom test
6. Look into the possibility of contrast reversal tests.

Future work should include reduction of data on three other systems which have been tested on the zoom optical target simulators. These are Martin-Marietta, McDonnel-Douglas, and a second Honeywell. All of these systems use a CCD or CID and have higher density than the 50 X 50 RETICON. These also are representative of the direction that the array technology is heading. At this stage, the CID with random access appears to be a leading candidate for EO trackers in that partial scene processing helps alleviate some of the rather significant scene processing problems and permits higher scan rates. This device is not here yet, but its potential is exciting.

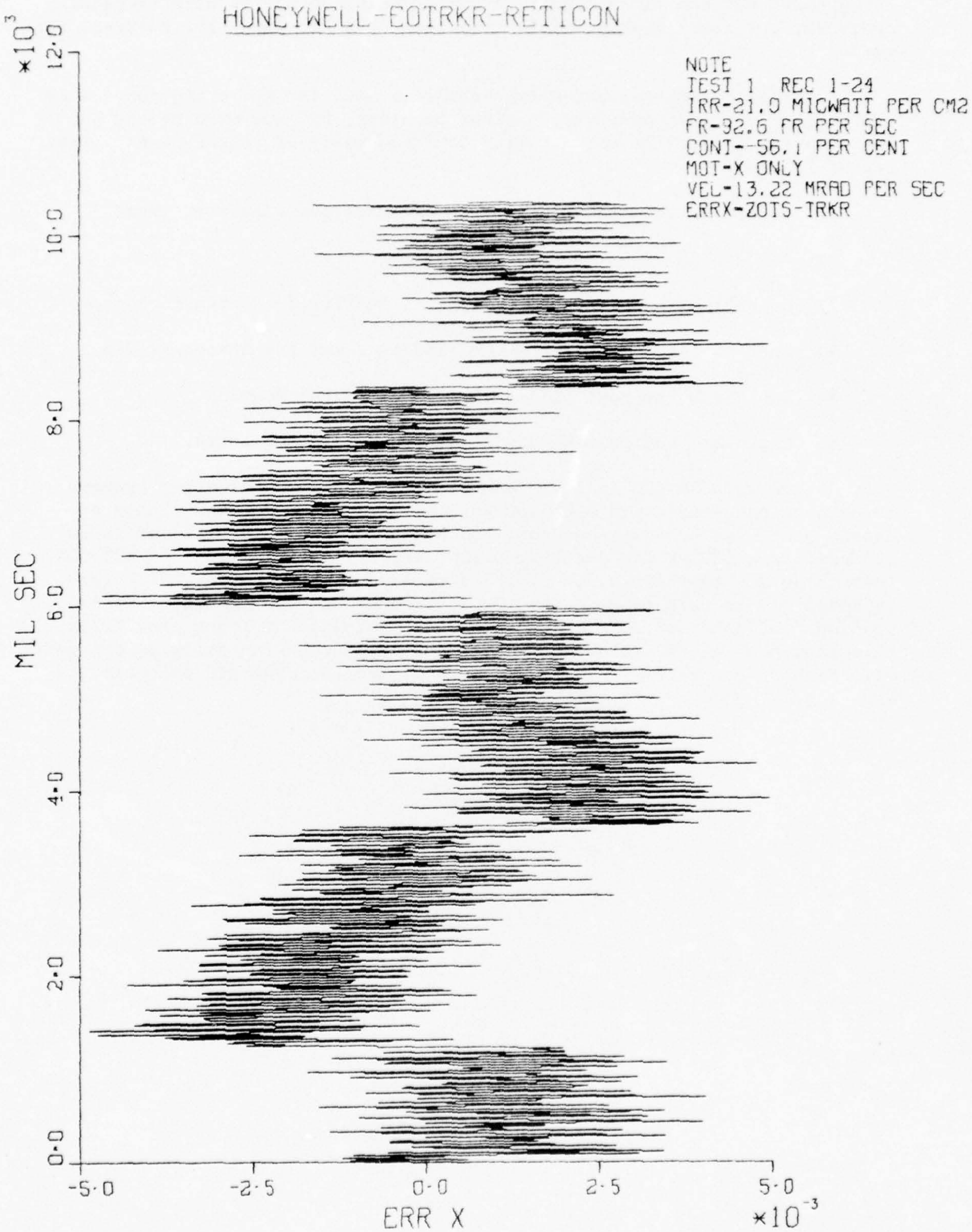


Figure 1. Tracking Error Vs Time

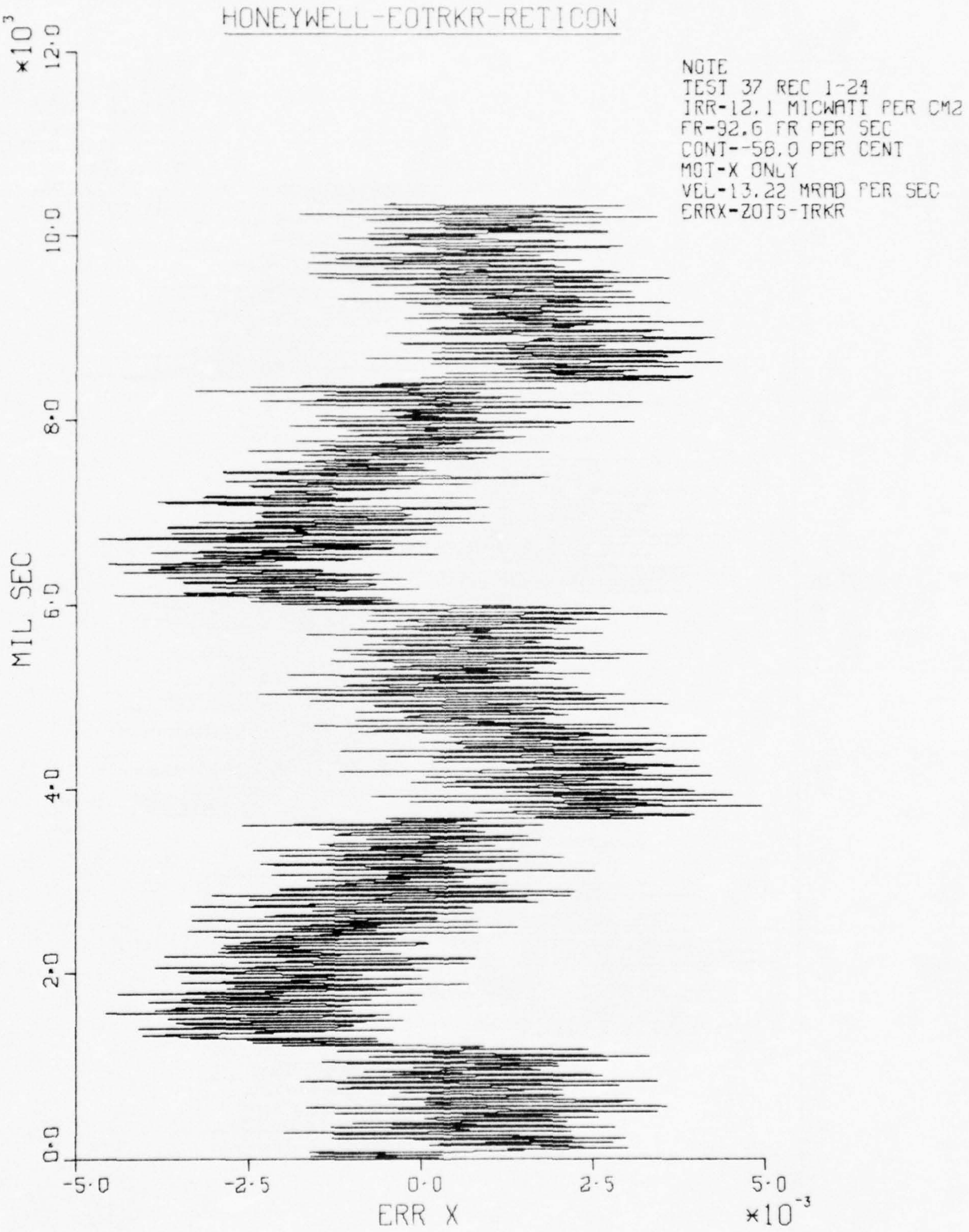
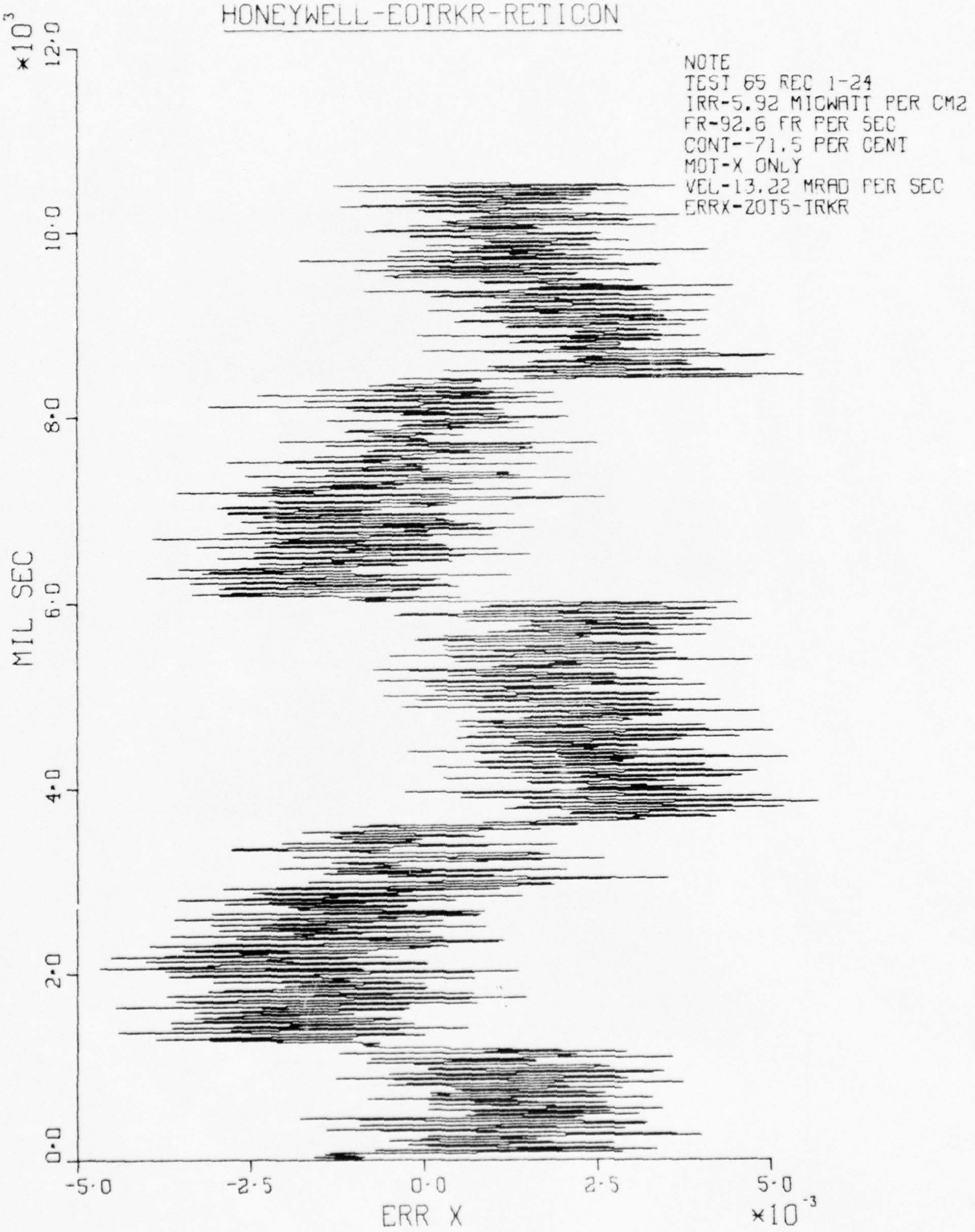


Figure 2. Tracking Error Vs Time



NOTE
TESI 65 REC 1-24
IRR-5.92 MICWATT PER CM2
FR-92.6 FR PER SEC
CONT--71.5 PER CENT
MOT-X ONLY
VEL-13.22 MRAD PER SEC
ERRX-ZOTS-TRKR

Figure 3. Tracking Error Vs Time

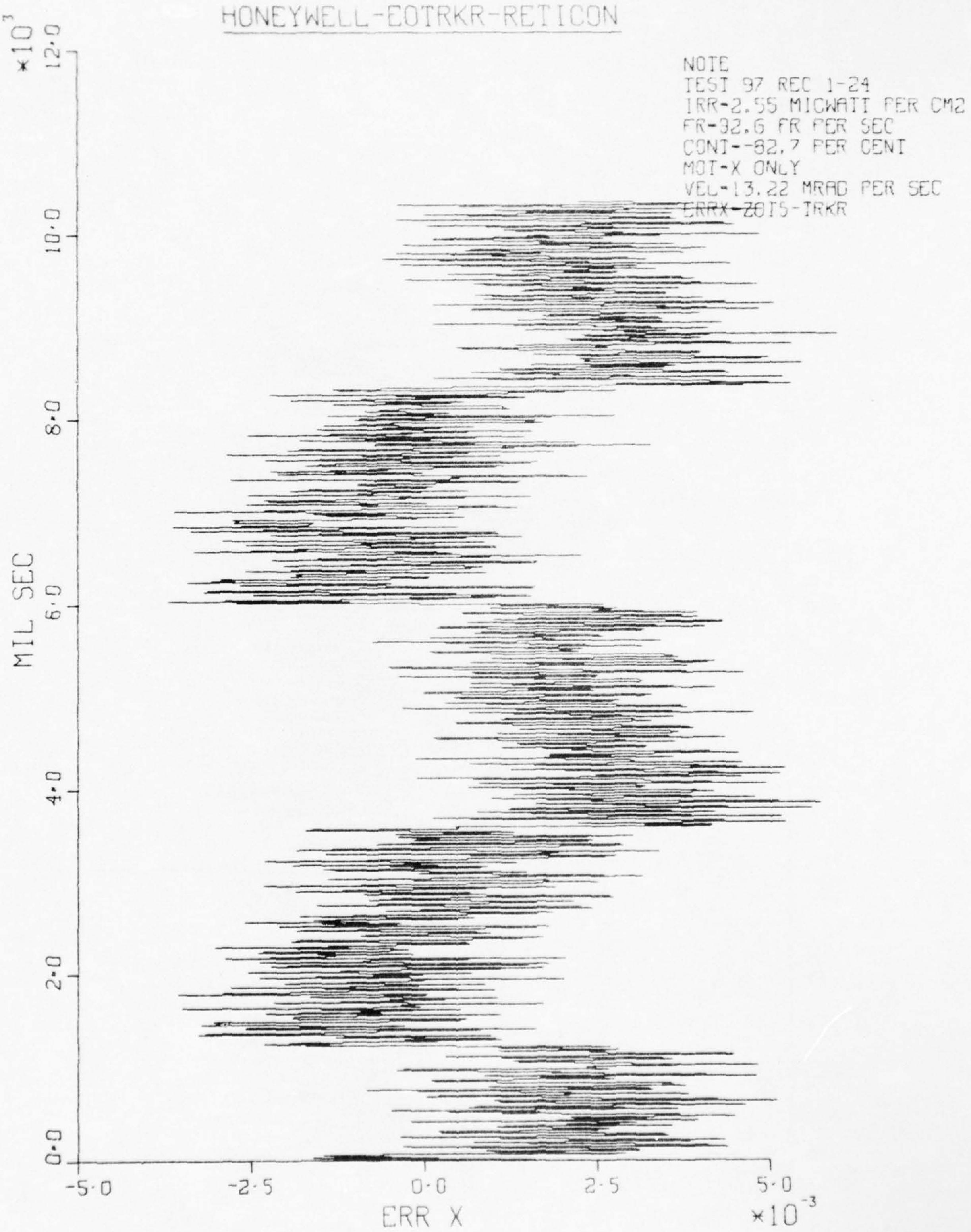


Figure 4. Tracking Error Vs Time

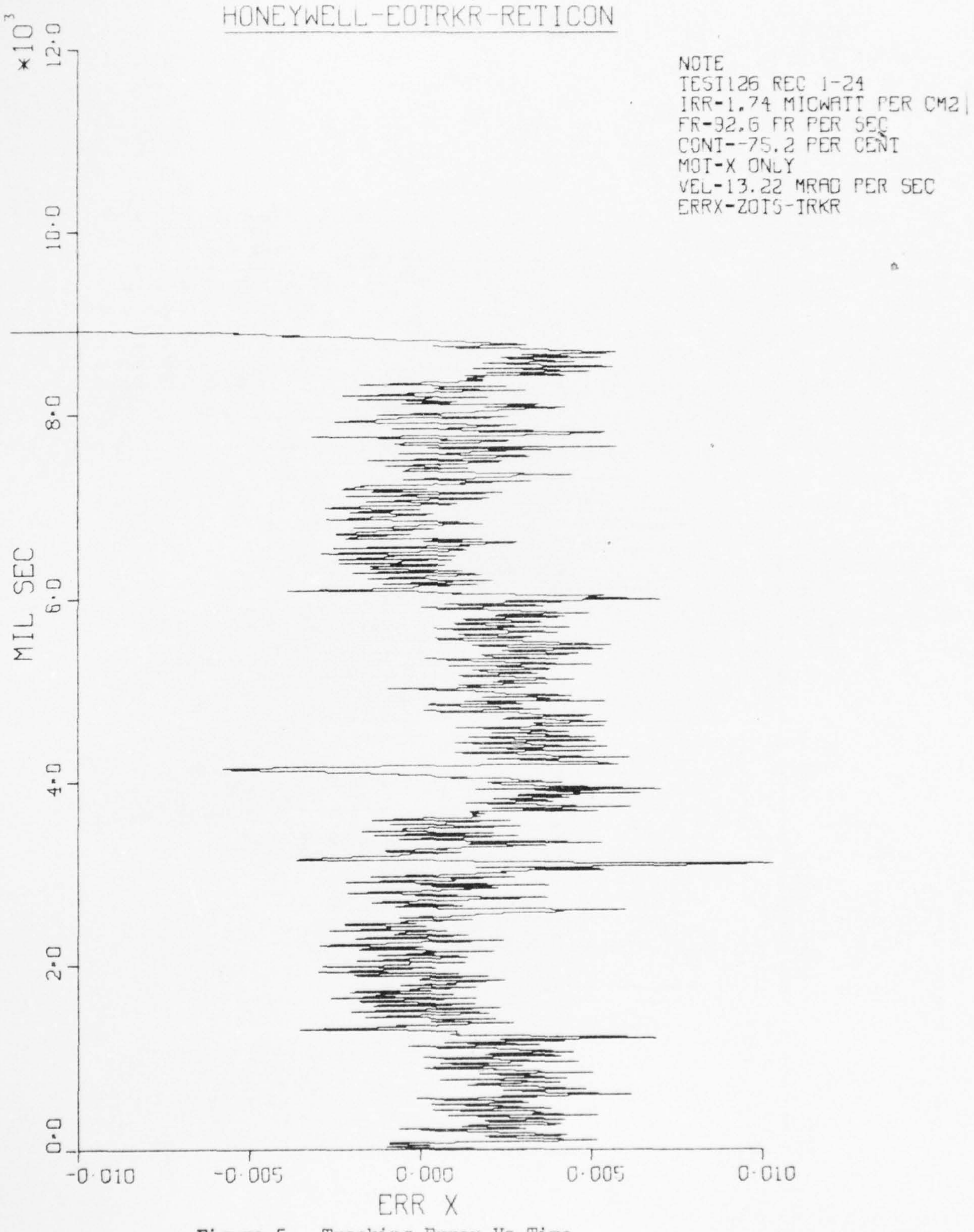


Figure 5. Tracking Error Vs Time

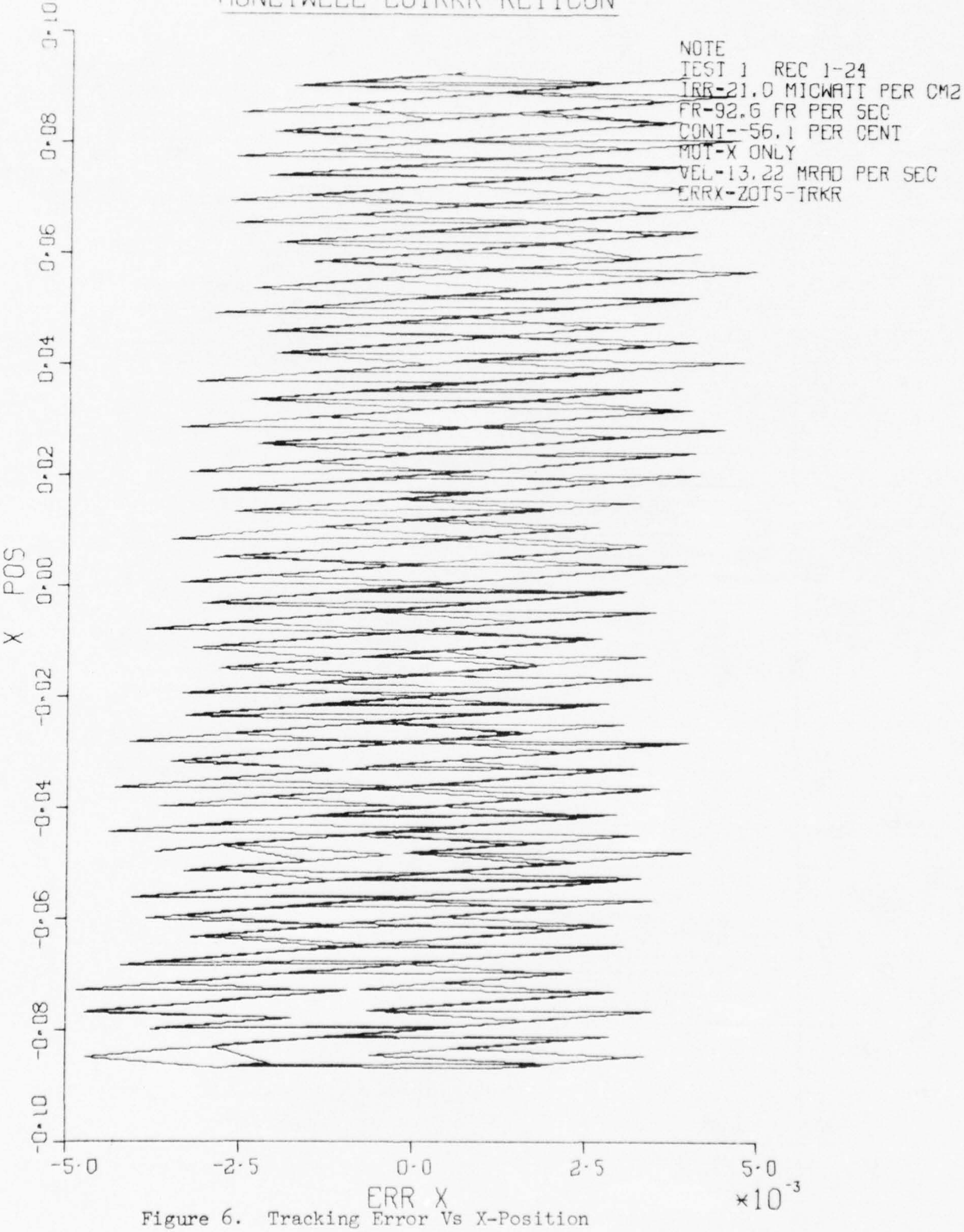


Figure 6. Tracking Error Vs X-Position

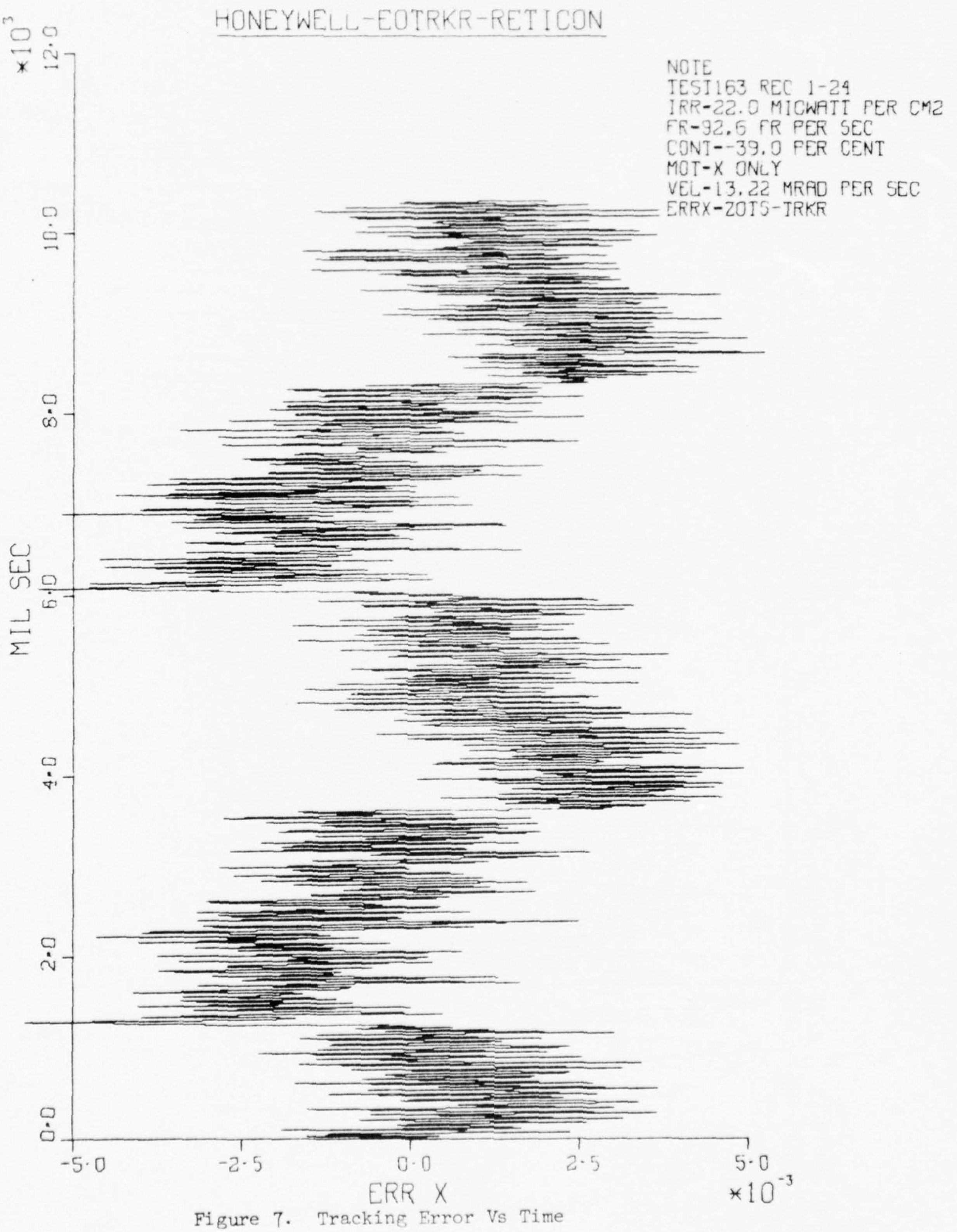


Figure 7. Tracking Error Vs Time

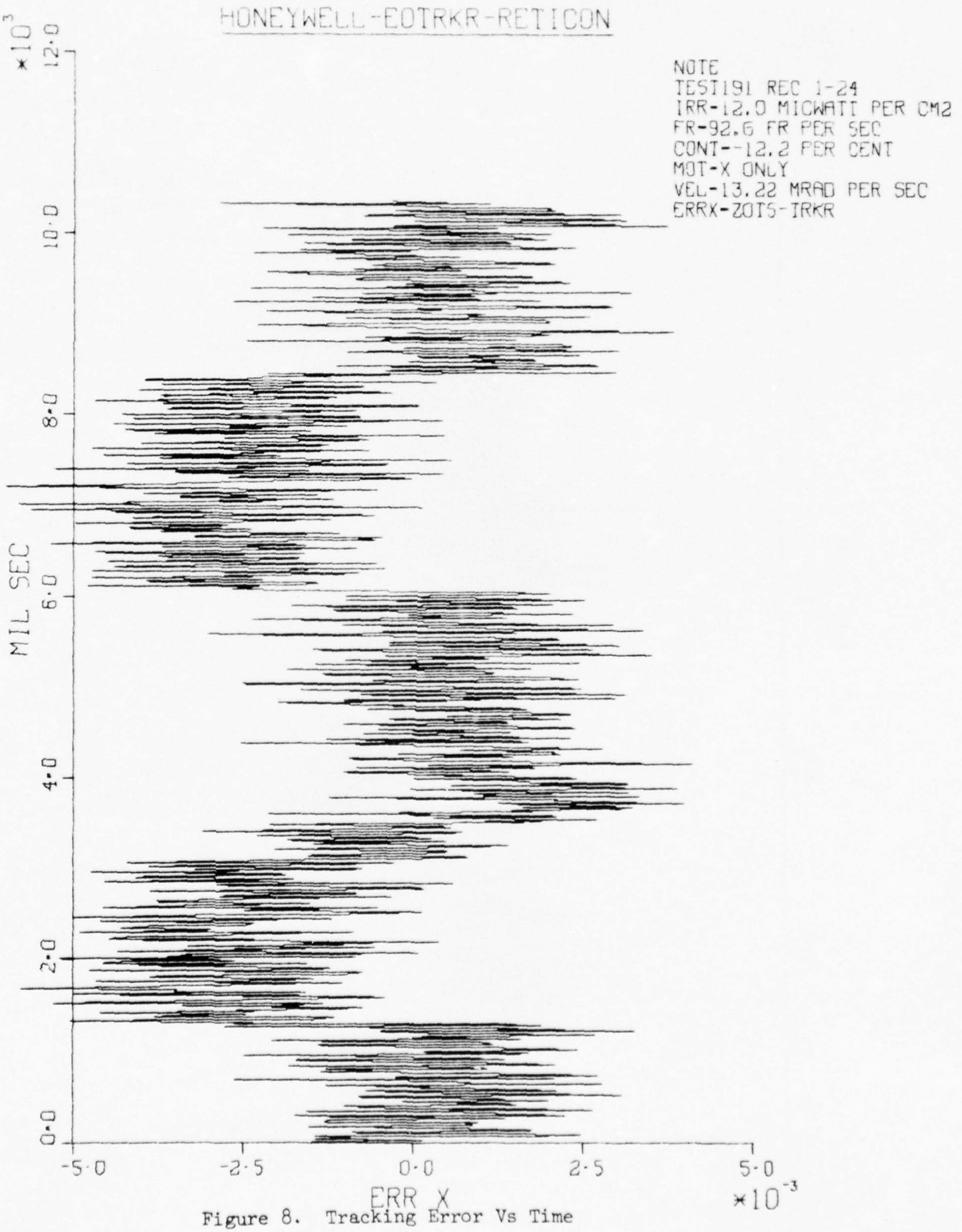


Figure 8. Tracking Error Vs Time

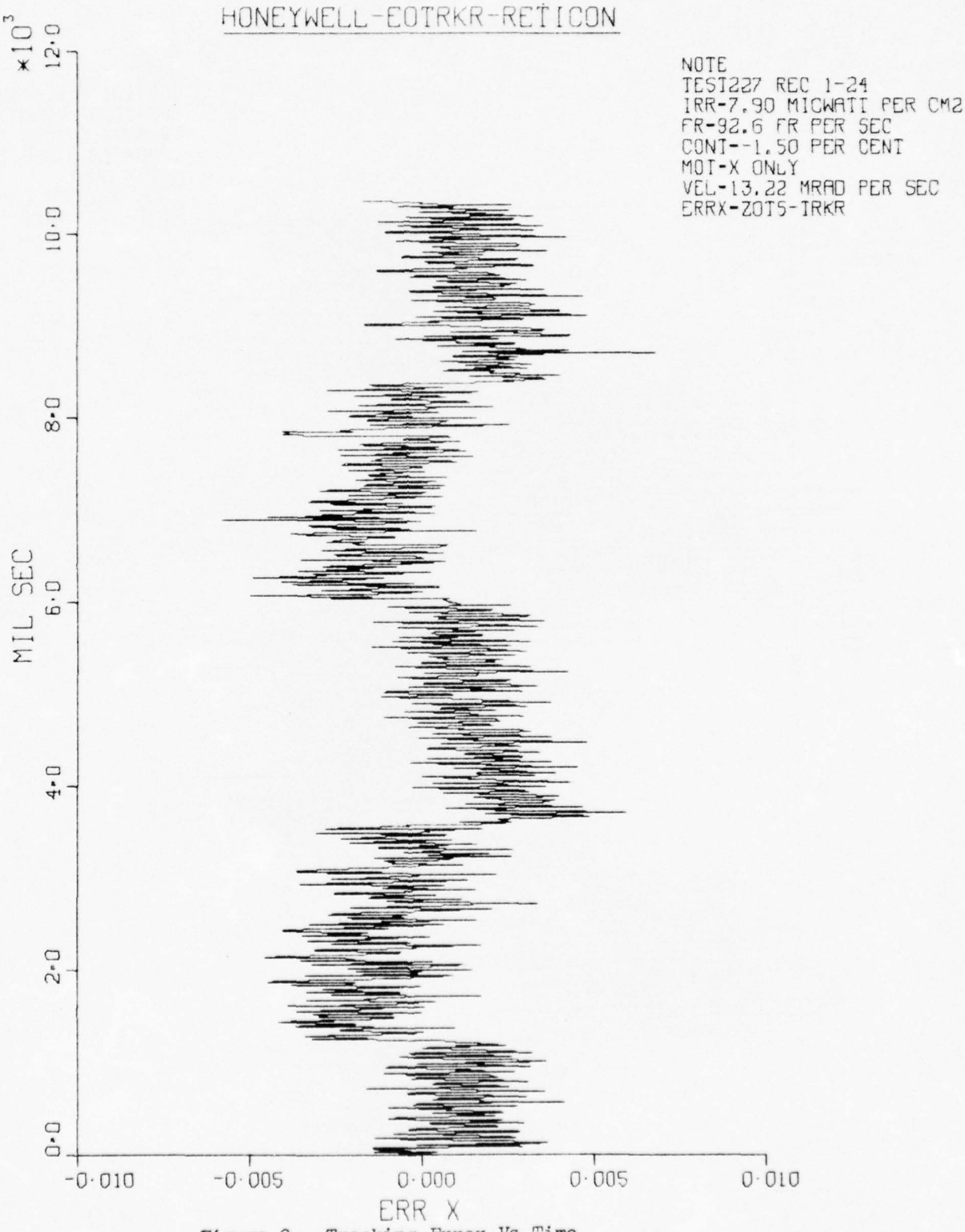


Figure 9. Tracking Error Vs Time

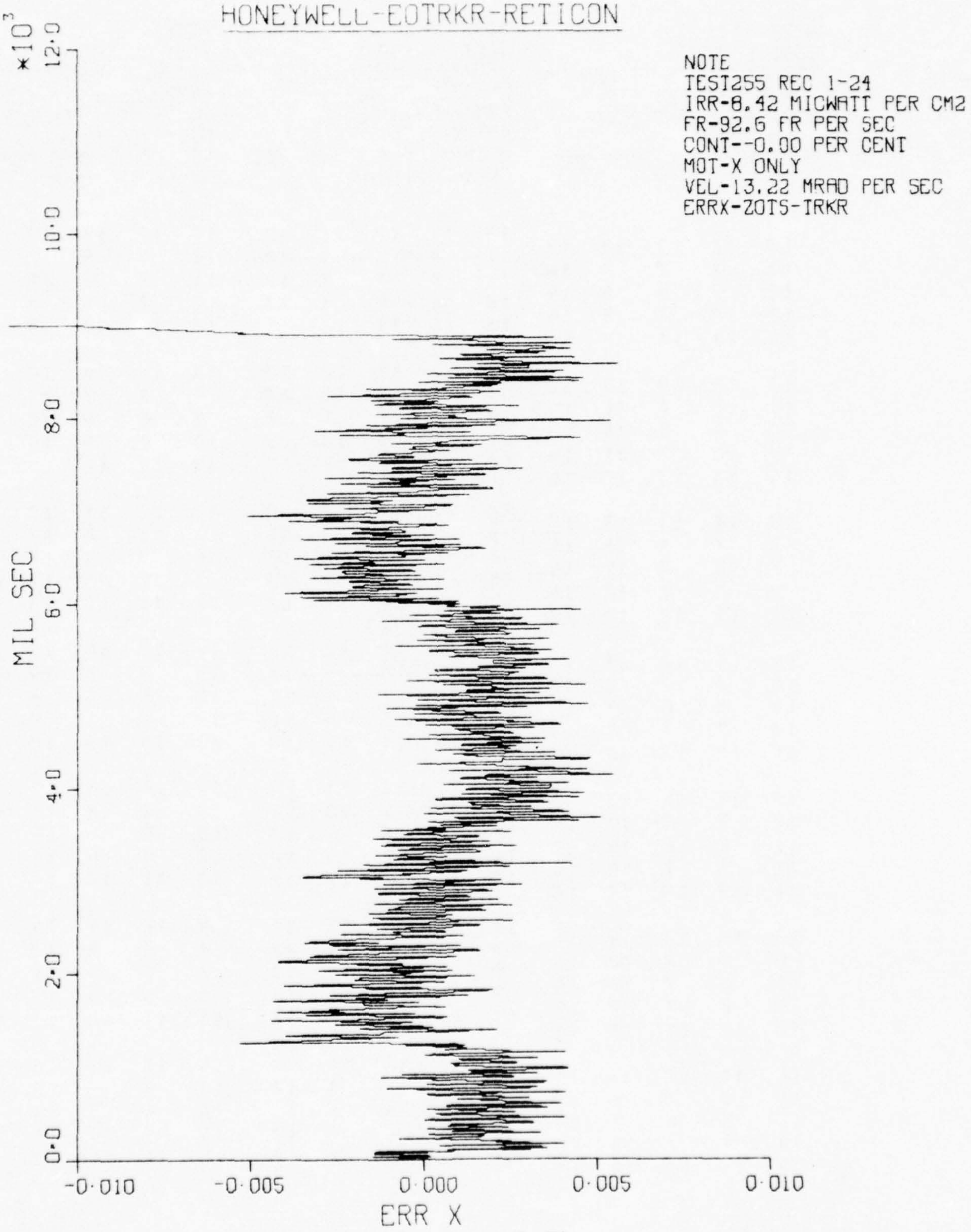


Figure 10. Tracking Error Vs Time

```
NOTE
TEST 1  REC 1
IRR-21.0 MICWATT PER CM2
FR-92.6 FR PER SEC
CONT--56.1 PER CENT
MOT-X ONLY
VEL-13.22 MRAD PER SEC
ERRX-ZOTS-TRKR

 13   13   13   13   13
 12    7    7    9   14
 12    6    6    8   14
 12    7    7    8   14
 14   13   13   13   14

 13   12   12   13   13
 10    7    7   11   14
 10    6    6   10   14
 10    8    7   11   14
 14   13   13   13   14

 13   13   13   13   13
  8    7    8   13   14
  7    6    7   12   14
  8    7    7   12   14
 13   13   12   13   13

 14   13   12   13   13
 12    7    7    8   13
 12    6    6    8   14
 12    7    7    8   13
 14   13   13   13   13

 13   13   13   13   13
 11    7    7   10   14
 11    6    6   10   14
 11    7    7   10   14
 14   13   13   13   13

 13   13   13   14   13
  9    7    8   12   14
  8    6    7   12   14
  9    7    7   12   14
 13   13   13   13   13

 14   13   12   13   14
 14    8    7    8   13
 14    7    6    7   13
 14    8    7    8   13
 14   13   13   13   13
```

```
NOTE
TEST 37 REC 1
IRR-12.1 MICWATT PER CM2
FR-92.6 FR PER SEC
CONT--55.0 PER CENT
MOT-X ONLY
VEL-13.22 MRAD PER SEC
ERRX-ZOTS-TRKR

 13   11   12   12   12
 11    8    8    9   12
 11    7    7    7   12
 12    8    6    8   13
 13   11   12   12   12

 13   11   12   12   12
  9    8    8   10   13
 10    7    7    9   12
 11    8    7   10   13
 12   11   12   12   12

 12   11   12   12   12
  8    8    8   12   12
  8    7    7   11   13
  8    8    7   12   13
 12   11   12   12   12

 13   11   12   12   12
 11    9    9    9   12
 12    7    7    8   12
 12    7    6    8   12
 13   10   11   11   12

 13   11   12   13   12
 10    8    8   10   12
 11    7    7    9   13
 11    7    7   10   12
 13   11   12   12   12

 13   11   13   13   12
  9    8    8   11   12
  9    6    7   11   13
  9    7    7   11   13
 12   10   12   12   13

 14   12   12   13   13
 12    9    8    8   12
 14    8    6    6   12
 14    8    6    8   12
 14   11   12   12   12
```

```
NOTE
TEST 65 REC 1
IRR-5.92 MICWATT PER CM2
FR-92.6 FR PER SEC
CONT- 71.5 PER CENT
MOT-X ONLY
VEL-13.22 MRAD PER SEC
ERRX-ZOTS-TRKR

 13   12   13   12   13
 10    7    6    8   13
 11    6    5    8   13
 12    7    5    8   13
 13   12   13   13   13

 13   12   13   12   13
  9    8    7   11   13
  9    6    6   10   13
 10    7    5   10   13
 13   11   12   12   13

 13   12   13   12   12
  7    8    8   12   13
  6    6    6   12   13
  7    6    6   12   13
 12   11   12   13   13

 14   12   14   13   13
 11    8    8   10   12
 12    5    6    8   13
 12    6    6    9   13
 13   10   12   12   13

 14   12   13   14   12
 10    8    8   11   13
 10    5    6   10   13
 10    6    6   10   13
 13   11   12   12   13

 14   12   14   14   13
  8    7    8   12   13
  8    5    6   11   13
  8    6    6   12   13
 13   11   12   13   13

 14   13   13   14   14
 13    8    8    9   13
 14    6    5    7   13
 14    6    5    9   13
 14   11   11   12   13
```

Figure 11. Array of Video Tracking Integers

```
NOTE                             NOTE                             NOTE
TEST 97 REC 1                    TEST125 REC 1                    TEST163 REC 1
IRR-2.55 MICWATT PER CM2         IRR-1.73 MICWATT PER CM2         IRR-22.0 MICWATT PER CM2
FR-92.6 FR PER SEC               FR-92.6 FR PER SEC               FR-92.6 FR PER SEC
CONT--82.7 PER CENT              CONT--75.2 PER CENT              CONT--39.0 PER CENT
MOT-X ONLY                       MOT-X ONLY                       MOT-X ONLY
VEL-13.22 MRAD PER SEC           VEL-13.22 MRAD PER SEC           VEL-13.22 MRAD PER SEC
~~ERRX-Z~~OTS-TRKR               ERRX-ZOTS-TRKR                   ERRX-ZOTS-TRKR

12  12  13  12  12               12  13  13  12  11               14  13  14  14  14
 8   7   7   9  12                8  11   9  11  12               11  10  10  14  13
10   7   6   8  12                8   8   6   9  11               12  10  10  13  14
11   9   6   9  12               10  10   6   9  12               13  11  10  14  14
12  13  12  12  12               11  12  10  12  12               14  13  14  14  14

12  12  13  12  12               12  13  12  11  12               15  13  14  14  14
 7   8   7  11  12                7  10   8  11  11               14  11  10  11  13
 7   7   6  11  12                7   9   7  11  12               15  11  10   9  14
 8   7   6  11  12                8  10   7  10  12               14  12  10  11  14
12  12  12  12  12               10  13  11  12  11               15  13  14  14  14

13  11  13  13  12               12  14  14  12  11               15  13  14  14  14
11   7   8   9  12                7   9   9  11  10               14  11  10  11  13
12   6   6   7  13                6   8   8  12  12               15  10  10   9  14
12   6   6   7  12                7   9   7  12  12               14  11  10  11  14
13  10  12  12  12                9  11  10  12  11               15  13  14  14  14

13  11  13  13  12               14  11  13  13  12               15  13  14  14  14
 9   7   8  10  12               10   8  11  11  11               13  11  10  12  13
11   6   6   9  13               10   5   9  10  13               14  10  10  10  14
10   6   7   9  13               10   7   8  10  12               14  11  10  12  14
13  11  13  12  12               13   7  12  11  12               15  13  14  14  14

13  11  14  13  12               15  13  12  14  13               14  13  14  14  14
 8   7   8  11  12               13  10   8  10  12               13  10  10  13  13
 9   5   6  10  13               15   9   5   7  11               14  10  10  11  14
 9   6   7  11  12               15   9   6   9  12               14  11  10  13  14
12  10  12  12  12               15  11   9  12  12               15  13  14  14  14

15  13  12  13  14               15  12  12  14  13               14  13  14  14  14
14   8   6   7  12               14  11   8  11  12               11  10  10  14  13
14   7   5   6  12               14   8   6   8  12               12  10  10  12  14
14   8   5   8  12               15   8   6  10  12               13  11  10  14  14
15  12  11  13  12               15  11   9  13  13               14  13  14  14  14

14  12  12  14  14               15  13  12  14  13               15  13  14  14  14
13   7   7   9  14               13   9   8  10  12               14  12  10  10  13
14   6   5   8  13               15   7   5   9  12               15  12  10   9  14
14   7   5   9  14               14   7   5  10  12               15  12  10  11  14
15  11  11  13  13               15  11   9  12  12               15  14  14  14  14
```

Figure 12. Array of Video Tracking Integers

```
NOTE
TEST191 REC 1
IRR-12.0 MICWATT PER CM2
FR-92.6 FR PER SEC
CONT--12.2 PER CENT
MOT-X ONLY
VEL-13.22 MRAD PER SEC
ERRX-ZOTS-IRKR

 15   15   15   15   15
 13   11   11   14   14
 13   10   10   13   15
 14   12   11   14   15
 15   15   15   15   15

 15   15   15   15   15
 11   12   12   14   14
 12   10   10   14   15
 12   11   10   15   15
 15   15   15   15   15

 15   15   15   15   15
 14   12   11   12   14
 15   11   10   10   15
 15   12   11   12   15
 15   15   15   15   15

 15   15   15   15   15
 14   12   11   12   14
 15   11   10   11   15
 15   12   10   12   15
 15   15   15   15   15

 15   15   15   15   15
 14   11   11   14   14
 15   10   10   12   15
 15   11   10   14   15
 15   15   15   15   15

 15   15   15   15   15
 13   11   12   15   14
 14   10   10   13   15
 14   11   10   15   15
 15   15   15   15   15

 15   15   15   15   15
 12   12   12   15   14
 12    9   11   14   15
 12   11   10   15   15
 15   15   15   15   15
```

```
NOTE
TEST227 REC 1
IRR-7.90 MICWATT PER CM2
FR-92.6 FR PER SEC
CONT--1.50 PER CENT
MOT-X ONLY
VEL-13.22 MRAD PER SEC
ERRX-ZOTS-IRKR

 14   13   14   12   14
 13   11   12   12   13
 14   11   11   11   14
 14   12   11   12   14
 14   13   14   13   14

 14   13   14   13   14
 13   12   12   12   13
 14   11   11   11   14
 14   12   11   12   14
 14   13   14   13   14

 14   13   14   13   14
 11   12   12   13   13
 12   11   12   12   14
 13   12   11   13   14
 14   13   14   13   14

 14   13   14   14   13
 13   12   12   12   12
 14   12   11   11   14
 14   12   11   12   13
 14   13   14   14   13

 15   13   14   14   13
 13   12   12   13   12
 14   11   12   11   14
 14   12   11   13   13
 14   13   14   14   14

 14   13   14   14   13
 12   12   11   14   12
 13   11   11   12   14
 13   12   11   14   13
 14   13   14   14   14

 15   13   14   14   15
 14   12   12   12   13
 15   12   11   10   14
 15   12   11   12   14
 15   13   14   14   14
```

```
NOTE
TEST255 REC 1
IRR-8.42 MICWATT PER CM2
FR-92.6 FR PER SEC
CONT--0.00 PER CENT
MOT-X ONLY
VEL-13.22 MRAD PER SEC
ERRX-ZOTS-IRKR

 14   13   14   13   14
 13   12   12   12   13
 13   12   11   11   14
 14   12   11   12   14
 14   13   14   13   14

 14   13   14   13   14
 12   12   12   13   13
 12   12   12   12   14
 13   12   11   12   14
 14   13   14   13   14

 14   13   14   13   14
 12   12   12   13   13
 12   12   12   12   14
 12   12   11   13   14
 14   13   13   13   14

 14   13   14   13   14
 11   12   12   13   13
 12   12   12   13   14
 12   12   11   13   14
 14   13   14   13   14

 15   13   14   14   13
 13   12   12   13   12
 14   12   12   12   14
 14   12   11   13   13
 14   13   14   14   13

 15   13   14   14   13
 13   12   12   14   13
 12   11   12   13   14
 12   12   11   14   14
 14   12   14   14   14

 15   13   14   14   15
 14   13   12   12   13
 15   12   11   11   14
 15   12   11   12   14
 15   13   14   14   14
```

Figure 13. Array of Video Tracking Integers

APPENDIX A

Zoom Optical Target Simulator (ZOTS)*

The Zoom Optical Target Simulator, shown schematically in A-1, projects the image of a target scene which can be presented to the input optics of a terminal guidance seeker system. The target scene can be expanded in size to simulate closure, varied in brightness and contrast, and moved to simulate target motion. The image can be presented in two ways:

1. Direct Mode: The image is projected directly through a collimating lens with a 3.15° FOV to the seeker.
2. Projection Mode: The image is projected onto a translucent screen, and the seeker is set up to view the screen through a collimating lens.

The simulator has a zoom ratio greater than 100:1 and can simulate a closure rate of more than 2,000 ft/sec. Zooming is done by moving both the carriage assembly holding the target plate, and the macro lens. Maximum slant range to target at the beginning of the simulated trajectory, and minimum slant range at simulator cut-off is a trade-off based on the maximum zoom ratio and the scale of the target scene being projected. The launch envelopes of all known video seeker systems can be accommodated with the simulator.

The target motion other than zoom is accomplished by moving the target plate. A rotation of the plate holder causes an image roll of up to ±15°. Translation motion in two axes of up to ±1 inch is possible. The amount of target movement that the seeker sees is dependent on the zoom position. All plate holder motion is servo controlled.

The light source is a 2,000 watt Tungsten-Halogen lamp. Energy from this lamp is focused on a condensor lens which cuts off radiation above 0.7 microns. This reduces heating of the photographic plate. The image is then focused and projected through the macro lens, collimating or projection lenses, picture wedge assembly and color correction filter, to the system under test. Image brightness is controlled with the picture wedge assembly, and opposed pair of neutral density wedges with >500:1 attenuation, and additional filters if needed. The ZOTS is designed to provide an image with a luminance of 0 to 180 foot lamberts in the projected

*Extracted from the: Central Inertial Guidance Test Facility, Laboratory Application Brochure, 6585th Test Group/GDL, Holloman AFB, New Mexico, April 1, 1974.

mode and 0 to 60,000 foot lamberts in the direct mode.

Another important effect is simulated with ZOTS. This is the back-scattering of radiant flux from the intervening atmsophere. The target picture is always illuminated with a constant intensity light source. The radiance of the projected scene is controlled with the servo-operated linear picture wedge assembly. Back-scattering is simulated by adding radiance to the projected scene. This is done with a separate light source with its own servo-controlled linear wedge assembly, designated the "Eogger". In the direct mode, the energy from the fogger unit is added with a beam splitter, and color corrected to 5900° K. In the projection mode, the fogger unit is mounted on the side of the projection lens, and the back-scatter radiance is superimposed on the target image at the screen. Color correction is provided by the screen in this mode.

Optical quality is exceptionally good. Resolution is 15 micro-radians over the zoom range, except at very close range. This is equivalent to a 1.5 ft object at 100,000 ft, or a 0.9 inch object at 5,000 ft.

The ZOTS also has the capability to simulate a target (i.e., a tank or truck) moving through a background scene. -This is done by moving the photographic image of the target on a film overlay across the plate holding the background scene.

Simulator functions of zoom rate, zoom position, picture wedge position, fogger wedge position, target translation and target rotation can all be controlled at the simulator panel, or by analog signal inputs. With this versatility, the system can be used to test a seeker system to determine the effect of a change in one or more of the above conditions. It can be used most effectively in closed loop simulation to simulate the changing conditions that a seeker system experiences when acquiring, tracking, and during a simulated flight trajectory. During this mode of operation the hybrid computer is used to drive the simulator as in the HITS simulations, and the three-axis flight simulator or a single axis rate table is used to dynamically position the seeker or guidance unit.

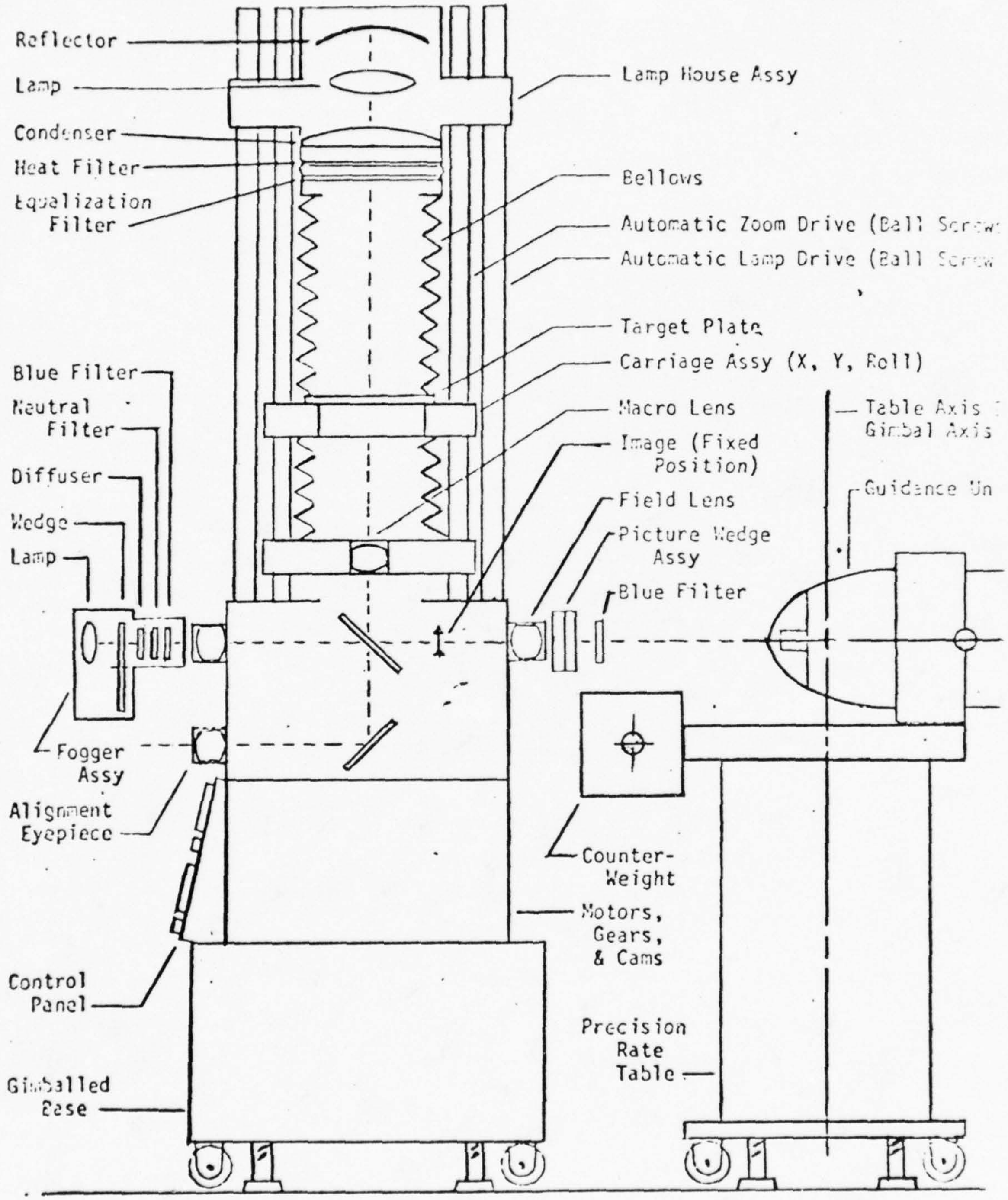


Figure A-1. Schematic of ZOTS – Zoom Optical Target Simulator

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

A REAL TIME TERMINAL GUIDANCE SIMULATION FACILITY

Prepared by:	Richard J. Wolf, Phd.
Academic Rank:	Assistant Professor
Department and University:	Department of Industrial Engineering New Mexico State University
Assignment:	
(Laboratory)	Avionics
(Division)	Electronic Warfare
(Branch)	Active ECM
USAF Research Colleague:	E. F. Mayleben
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

A REAL-TIME TERMINAL GUIDANCE SIMULATION FACILITY

By

Richard J. Wolf

ABSTRACT

The Avionics Laboratory of Wright-Patterson Air Force Base is providing a real-time, hardware-in-the-loop, missile simulation capability to its Anechoic Chamber so that a more realistic and valid investigation of electronic countermeasures to terminal homing RF guided missiles will be possible. This simulation facility will provide closed loop simulation. Such a simulation facility requires a computer hardware and software combination to simulate the missile flight dynamics and missile-to-target relative geometry.

This report discusses the design of this computer system. The design of the entire simulation facility is discussed inasmuch as this is related to the computer system design. The criteria for choice of the computer hardware is presented. The recommended approach toward software development is outlined and the requirements of the simulation computer are discussed.

ACKNOWLEDGMENT

The author is sincerely grateful to the Air Force Systems Command for support of this summer research. A deep appreciation is due to ASEE and Auburn University and in particular Mr. Fred O'Brien and Major David E. England for excellent administration of the project. Their work has certainly made this summer more productive and enjoyable.

The author is indebted to the Active ECM Branch of the Electronic Warfare Division of the Avionics Laboratory for providing a cordial and helpful work environment. It has been a pleasure and a professionally gratifying experience to work with these very capable people. In particular, thanks is due to Mr. James Kastle, Mr. E. F. Mayleben, Mr. H. Carlstrom, Mr. Bill Clark, and Ms. Marge Zampatti.

It has been enjoyable and profitable in working with the staff of IITRI, namely Mr. R. R. Rathbun and Mr. R. Rhodes. I wish to express my thanks to them.

INTRODUCTION

The Avionics Laboratory of Wright-Patterson Air Force Base is providing a real-time, hardware-in-the-loop, missile simulation capability to its Anechoic Chamber so that a more realistic and valid investigation of electronic countermeasures to terminal homing RF guided missiles will be possible. This simulation facility will provide closed loop simulation thereby investigating the combined effect of electronic countermeasures, simulated missile flight dynamics, system non-linearities, and hardware imperfections. Such a simulation facility requires a computer hardware and software combination to simulate the missile flight dynamics, missile-to-target relative geometry, and in some cases, the missile autopilot. This combination is interfaced with missile hardware including the missile seeker and guidance computer, and, if available, the missile autopilot. In addition, an interface to the target angular positioning system is required so that the RF energy is radiated from the correct target-to-missile relative angle.

The design of such a missile simulation facility requires appropriate consideration of the trade-offs between computational methods, hardware complexity, software complexity and maintainability, cost, and simulation validity and applicability. The computation may be accomplished by analog, hybrid, or general purpose digital computers. Within the class of general purpose digital computers, there are many architectures at varying costs which are especially adaptable to various types of computational problems. Each general purpose computer system includes software development and support aspects which have an important influence on the simulation system life cycle cost. General purpose computer systems are especially adaptable to applications secondary to the simulation system and the utility of these applications should be considered in the simulation design.

As the simulation generality increases, the system cost increases. Identification of immediate and foreseeable goals is, therefore, important to the simulation system design.

This report documents efforts involved in the design of the terminal homing missile simulation facility computational elements. These efforts necessarily involve consideration of the missile hardware interface and the target angular positioning system, but do not include the design of these simulation system components.

REAL-TIME SIMULATION

Real-time, closed-loop simulation enhances considerably the validity and utility of a simulation effort as compared to that of analytical simulation (i.e., one that involves no missile hardware). This is due to the fact that actual hardware is involved with simulated flight dynamics and simulated effective target signature. In other words, one need not concern himself with hypothesizing the hardware performance in the untested countermeasure environment or assuming that certain aspects of performance will remain unchanged from other tested environments. The analyst is merely required to observe the hardware performance.

Terminal homing missiles of the surface-to-air class utilize the principle of proportional navigation for guidance (with possibly some exception). Under this principle the missile autopilot at any instant attempts to maintain a constant angular change between the missile velocity vector and the line of sight to the target. This can be expressed as requiring the missile velocity vector time-rate-of-change to be proportional to the missile to target line of sight time-rate-of-change where the proportionality constant (navigation constant) may be fixed or may change with time.

In implementing this guidance principle, the RF seeker antenna attempts to center the incoming radiation (the effective combination of target echo and countermeasures) by means of the head servo loop. The angular rate of any required antenna movement is then presented to the autopilot as the angular velocity of the missile attitude relative to the line of sight. The autopilot then utilizes feedback from the missile dynamics to compute the net angular rate of change of missile velocity vector relative to line of sight. Divergence of this rate from the desired rate is the tracking error and appropriate commands to the missile control surfaces are intended to drive this error to zero.

Some missiles use missile movement compensation or stabilization networks. With this system, sensor units detect missile movement and compensating movement is injected into the head servo loop so that the only antenna movement relative to inertial space is that due to line of sight rotation. Thus the rotation required to stabilize the antenna head represents missile attitude rotation relative to line of sight.

With this guidance concept in mind, reference to Fig. 1 will assist in outlining the concept of the real-time simulation facility. Fig. 1 identifies the general equipment configuration of the facility. In Fig. 1 the missile seeker is seen situated in a porthole in the Anechoic Chamber wall. The missile axis is aligned with the missile seeker axis (not the antenna axis) and projects toward the center of the Anechoic Chamber.

The missile seeker antenna axis, however, is in general not aligned with the missile body axis. This reflects the "lead angle" between the missile attitude or body axis and the line of sight to the target. This angle is the angle referred to as β in Fig. 1. Although not shown in Fig. 1, the angular position of the seeker antenna axis may not be the same as the angle β. This is because countermeasures and other effects may introduce a tracking error not discernible by the missile hardware.

If the missile seeker antenna axis is denoted as $\tilde{\beta}$, one can then refer to $\tilde{\beta}$ as the apparent angle between missile attitude and missile to target line of sight. The missile autopilot will obtain $\tilde{\beta}$ or perhaps the time-rate-of-change of $\tilde{\beta}$ and, by knowing the time-rate-of-change of missile attitude to velocity vector angle, will generate appropriate control surface commands so that the time-rate-of-change of $\tilde{\beta}$ will be driven to the value dictated by the proportional guidance law.

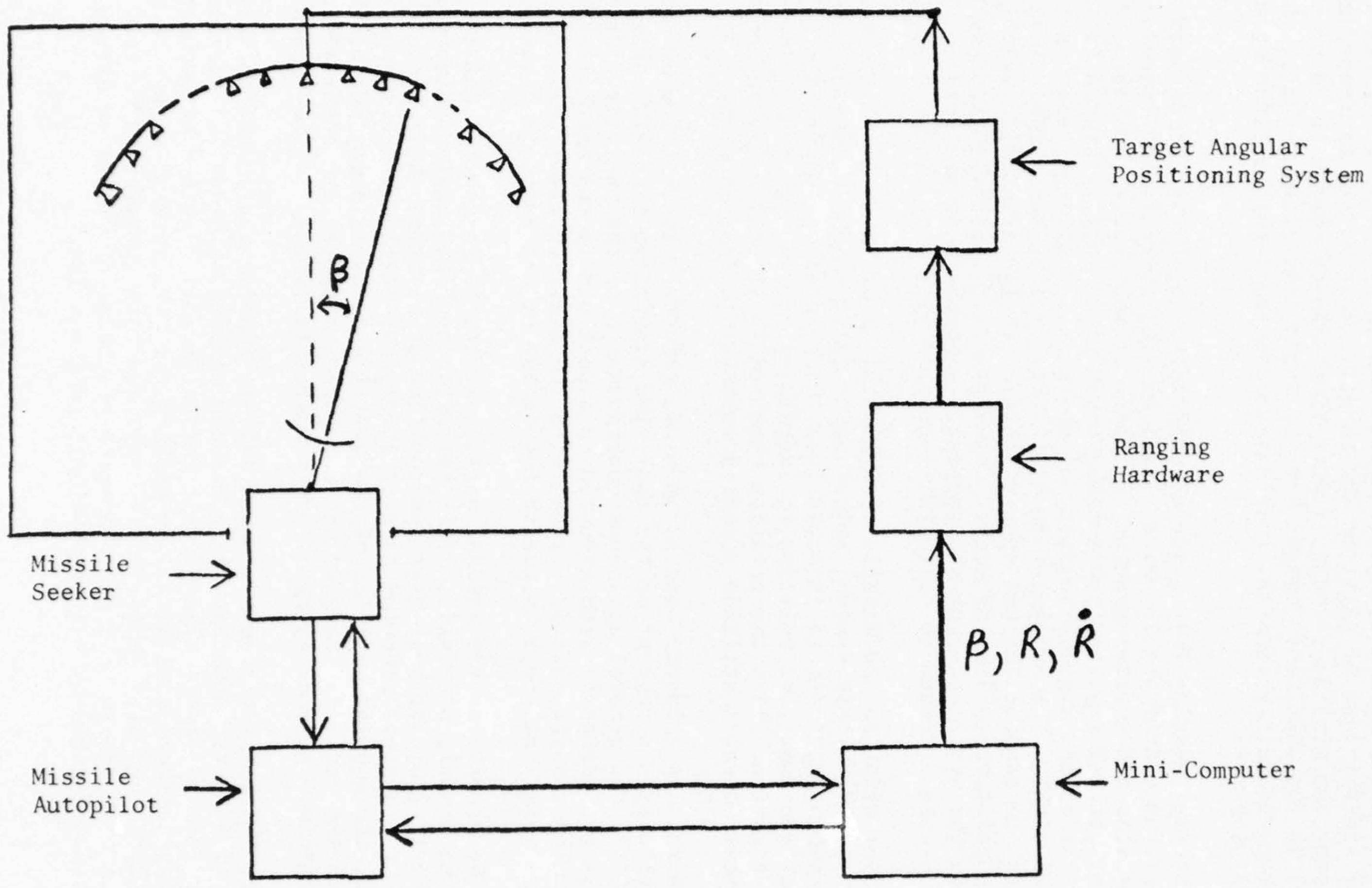


FIGURE 1. EQUIPMENT CONFIGURATION

The digital mini-computer will utilize geometric and kinematic software modules and these control surface commands to generate the trajectory of the missile. The target trajectory will be given as an input curve so that the true angle β may be calculated during each successive update interval. This angle, together with the range R and time-rate-of-change of range $\dot{R}$, will be passed to the target angular positioning system (which includes for purposes of this report, the range simulating hardware) so that the simulated target echo may be generated.

Fig. 1 and this discussion assume that the missile autopilot is present. If this is not the case, the autopilot must be simulated by the digital computer.

Because the missile dynamic motion relative to inertial reference cannot be effected within the Anechoic Chamber, the missile inertial reference sensors must be simulated by the digital computer. This information must be translated to an appropriate signal and presented to the missile hardware. In the same manner control surface feedbacks must be simulated and presented to the missile hardware.

In Fig. 1 the target angular positioning system is represented as a linear array of transmitting antennas oriented on an arc of a circle centered at the seeker antenna. As will be discussed at a later time, this is not the only feasible approach to target echo simulation, but this method has been chosen for the Anechoic Chamber missile simulation facility. The choice of a linear array limits tracking errors to only one dimension as far as the missile hardware is concerned (one additional dimension is range which is not properly termed as tracking error, but is, however, a dimension of the simulation). This does not, however, limit target or missile motion to a horizontal plane; it merely necessitates the assumption of perfect tracking in one component of motion. The validity of this assumption depends upon the particular missile and countermeasures investigated and is an important consideration in the design and interpretation of each test utilizing the Anechoic Chamber simulation facility.

Based upon the above assumption, it is seen that the simulation software, to be completely general, must allow for three-dimensional target and missile trajectories and may simulate analytically the third dimension of tracking response.

As is typical with analytical simulations, the simulation facility must terminate the simulation at some point prior to what would be termination of the simulated engagement. This is true for two reasons. In the case of a miss, there is no point in continuing in simulating the engagement after a miss has been determined. In fact, the hardware arrangement in general prohibits this.

In the case of a successful engagement, the end game relative angles and angular rates become excessive. In the analytical case, error is introduced as these rates become large, and in the real-time simulation hardware, limitations may be reached in any or all of the aspects: angles, angular rates, or power required due to small ranges. This means that the real-time simulation must be terminated at some point prior to impact and the miss distance extrapolated from these final conditions.

In summary, the real-time hardware-in-the-loop simulation allows the missile hardware to receive the simulated target echo as modified by a countermeasure. The hardware interprets this signal as it would in an actual engagement and then computes appropriate guidance commands. These guidance commands are then involved, via digital simulation, with the aerodynamic characteristics of the missile and the missile kinematics to produce the effective geometrical relationship between the target and the missile. This information is then used to assist in providing target echo and countermeasure effect to the missile hardware. In this manner a very realistic representation of an actual engagement is possible.

TARGET ANGULAR POSITIONING SYSTEM

In the previous section reference was made to the target angular positioning system. For the purposes of this report, this system is assumed to include all of the electronics necessary to produce RF signals representing a given dynamic geometrical relationship of the missile-to-target engagement including presentation of electronic countermeasures. In addition to being important to the missile simulation in its own right, this system is important to the digital computer system design as it impacts on the computational requirements of the computer system.

There are several methods of simulating a given dynamic geometric relationship between the missile seeker and the target. In all methods considered the range relationship is represented by RF attenuation and the doppler effect. The angular relationship can be represented by at least four methods:

(1) A linear mechanical system,
(2) A linear antenna array,
(3) A two-dimensional antenna array, or
(4) A synthetic line-of-sight approach

For those seeker systems with missile motion compensation networks, a fifth method, mounting the seeker on a flight table, cannot represent target-to-seeker angular motion. This is because the motion compensation networks will maintain a constant seeker antenna to fixed target array relationship. In other words, in these systems, relative seeker-to-target angular motion must be produced by motion of the apparent target echo source. The incorporation of a seeker mounted on a flight table exercises the antenna stabilization network, but not the total tracking loop [6].

The linear mechanical system basically involves a single antenna mounted on a track. The track forms an arc of a circle centered at the seeker antenna. The transmitting antenna is propelled by an electric motor such that the desired angular relationship is portrayed. This method has been used successfully in other real-time simulation facilities [6]. It is relatively inexpensive and is easily maintained. The major problems with this method are the possible inaccuracy as compared with other methods, the limitation placed on angular accelerations due to the mass of the moving antenna, and possible phase distortions due to flexing of the coaxial cable input.

The linear antenna array consists of several potential transmitting antennas mounted on an arc of a circle centered at the seeker antenna. The centroid of the apparent target echo is moved electronically by varying the amplitude of the antennas. In practice only two adjacent antennas are transmitting at any instant. This method is probably significantly more expensive than the mechanical approach, but is definitely far less expensive than the third approach. This method suffers no practical limitations on angular accelerations and can be extremely accurate. It is less mobile than the first method in relation to movement in and out of the Anechoic Chamber. Both this method and this first method suffer a deficiency in allowing only one dimension of tracking error. This is considered to be a justified deficiency.

The third method, a two-dimensional antenna array, is by far the most generally applicable approach. It allows two dimensions of tracking error and is accurate and is not limited by angular accelerations. Offsetting these advantages is the tremendous cost of this system relative to either of the first two approaches. It would also be difficult to dismantle this system so that it could be removed from the Anechoic Chamber. It is also more difficult to calibrate this system.

The synthetic line-of-sight method has been used in real-time simulations [8]. The method essentially involves a single transmitting antenna which remains fixed in position. The relative angular movement is accomplished by injecting a bias into the seeker head servo loop such that the desired angle is obtained. This angle is then compensated before being presented to the autopilot. The method can be used to provide one or two dimensions of tracking error or can be used to augment methods one or two to provide two dimensions of tracking error. The method would result in increased software complexity for the digital simulation computer and would require missile hardware modification. Because of these factors the relative cost advantage or disadvantage of this approach is unclear. It is also difficult to substantiate the realism of simulation results produced by this method. It would perhaps be an attractive method of adding a second dimension to that of method two, but a detailed investigation is necessary before such a decision is made.

The second method, a linear antenna array, has been determined to provide the best potential results for the relative cost. The feasibility of adding a second dimension of tracking error through the synthetic line-of-sight technique should be investigated.

SIMULATION COMPUTER

Traditionally the computational requirements of a real-time hardware-in-the-loop simulation have been handled by analog computers. Analog computers are well adapted to the solution of differential equations such as those involved in this simulation. Analog computers are parallel processors and can match well with the simultaneous requirements of the missile hardware and the target echo presentation. Therefore, there is a substantial argument for the use of an analog computer in real-time missile simulation.

The disadvantages of an analog computer for this specific simulation application were such that a digital computer was deemed preferrable. Perhaps the most important criterion was that the staff of the Anechoic Chamber is familiar with digital computers to a much greater extent than with analog computers. This is seen to be especially important in the case of a simulation where, usually, the validity of results is difficult to establish. The advantage of confidence in the computations is very attractive in this case.

The relative cost of an analog computer with sufficient capacity for this application is difficult to determine without first producing a fairly detailed simulation model for the analog computer. It is, however, felt to be a safe assumption that the analog computer would cost more than the required digital computer including associated software development and support elements.

The adaptability of a digital computer to other tasks and projects has bearing on the relative system cost. The Anechoic Chamber facility presently has digital computation equipment which can be utilized to reduce the net cost of the digital computer required of the real-time simulation. The digital computer can be easily adapted to future Anechoic Chamber projects and can be easily and inexpensively upgraded to meet expanding requirements or changes in scope of the real-time simulation facility.

Once programmed, the digital computer requires no calibration, and hardware imperfections invariably lead to easily identifiable failure. In contrast, the analog computer must be continually calibrated and checked, for a slight maladjustment can lead to totally wrong, but seemingly correct results. This potential problem would diminish with experience, but does represent a disadvantage of the analog computer.

With these advantages taken into consideration, a digital computer approach was determined the better alternative. The advantages could be summarized as:

(1) Existing familiarity with digital methods
(2) Adaptability to other tasks and future expansion

(3) High reliability once programmed
(4) Utilization of existing digital resources
(5) Lower relative net cost

CHOICE OF DIGITAL COMPUTER

Modern technology has made available several computers of impressive computational capability which retain an attractive low cost. Choice among these computers (referred to as mini-computers) is nontrivial due to competitive pricing and overall attractiveness of computer architectures. Four computers were identified as being suitable for the present application.

Manufacturer support of the prospective mini-computer system was identified as a prime criterion. Thus, although other manufacturers and systems houses were contacted, no second attempt at contact was made if the vendor failed to respond. It was felt that this indicated lack of support for the product. In addition, the four mini-computers that were identified appeared to be very powerful and quite reasonable cost-wise.

As the simulation will utilize a substantial range of numbers, it was determined that floating point arithmetic would be essential. In some software tasks, such as numerical integration, double precision floating point will be required. Therefore, any prospective computer system must provide fast, efficient floating point hardware and must provide double precision instructions. Floating point operations were used as a primary criterion for the choice of the digital computer system.

Most machines which possess a fast floating point processor will also have a powerful standard instruction set and will execute standard instructions quickly. This was offered in support of the use of floating point processors as a criterion and was found to be essentially true.

It was soon found that, even for a very specific application, it is difficult to judge computers at machine level. Each of the four machines had particular qualities which were attractive over the other competitors, but there still seemed to be an overall balance among the machines. Some desirable attributes were identified and are listed in Table 1. These characteristics were used to make a preliminary ranking of the mini-computers as indicated in Table 1. Identification of the machines is given in Table 2.

The items listed in Table 1 are subjective characteristics in that they have been deemed "nice" to have, in general. The order of presentation of attributes in Table 1 is approximately the order of desirability, but this is only a preliminary and subjective ranking based upon considerations as listed in Table 1.

It is difficult to judge computers of a sophistication level such as these by comparing elements of respective instruction sets. The

<u>TABLE 1</u>

SPECIFIC COMPUTER ATTRIBUTES

ATTRIBUTE\COMPUTER NUMBER	I	II	III	IV
FLOATING POINT PROCESSOR SPEED	b	c	a	d
" " " PARALLEL OPERATION	Yes	Partial	No	No
" " " SUBJECTIVE USEABILITY	a	a	a	?
HARDWARE DOUBLE PRECISION	Yes	Yes	54 bit	Yes
GENERAL COMPUTATION SPEED				
Core Memory Cycle Time	600[1]	650[1]	660	500[2]
MOS Availability & Speed	~200[2]	450	330[2]	No
Bipolar Availability	No[3]	300	No	No
Memory Interleaving	Yes	Yes	Yes	Yes
GENERAL PURPOSE REGISTERS AVAILABLE TO ONE TASK	4	6	8	15
INDEXED ADDRESSING	Yes	Yes	Yes	Yes
AUTO INCREMENT	Yes	Yes	No	No
AUTO DEREMENT	Yes	Yes	No	No
HARDWARE STACK INSTRUCTIONS	Yes	Yes	No	Yes
OPERATING SYSTEM UTILITY FOR SOFTWARE DEVELOPMENT & IMPLEMENTATION	a	a	b	b

TABLE I (CONTINUED)

SPECIFIC COMPUTER ATTRIBUTES

ATTRIBUTE\COMPUTER	I	II	III	IV
EXPANDABILITY				
Multi-CPU with Shared Memory	No	Yes	Yes	Yes
Secondary General ADP Applications	a	a	b	--
Company History of Upward Compatibility	a	b	b	--
SUBJECTIVE OPINION OF ARRAY MANIPULATION	a	a	b	a
VECTORED INTERRUPT SYSTEM	H,S	H,S	H,S	H,S
MEMORY SIZE WITHOUT HARDWARE MAP	32K	28K	32K	32K
DIRECT MEMORY MANIPULATION INSTRUCTIONS	b	a	d	b
BYTE AND BIT ORIENTED INSTRUCTIONS	b	a	d	a
WORD SIZE	16	16	16	32
GENERAL IO CHARACTERISTICS	b	a	b	b
MULTI-LEVELS OF INDIRECTION	Yes	No	Yes	No
SOFTWARE SOPHISTICATION	a	b	b	b
TASK CONTEXT SWITCHING	b	b	d	a
WRITABLE CONTROL STORE	Yes	No	Yes	No

a, b, c, d = relative ranking

[1] with 4-way interleaving

[2] effective limit

[3] a 4-word cache of bipolar memory is provided

H = Hardware

S = Software

TABLE II

CANDIDATE COMPUTERS

I. DATA GENERAL ECLIPSE S/200

II. DIGITAL EQUIPMENT CORPORATION PDP 11/45

III. VARIAN DATA MACHINES V73

IV. MODULAR COMPUTER SYSTEMS MODCOMP IV

flexibility of assembler language programming allows several approaches to an identical problem and the net efficiency of an approach is context dependent. This means that one must be very careful when judging a computer based upon instructions and instruction times.

Because the one dominating requirement of the digital simulation computer is that it be fast enough to affect a real-time control, some means of judging the competing mini-computer's speed must be utilized. Toward this end a benchmark was constructed. The benchmark was written in FORTRAN. It incorporates arithmetic and logical operations similar to those in analytical simulation programs such as MICOM's 6 DOF Missile Simulation [1].

The benchmark basically incorporates three types of operations. The first is a two-dimensional floating point table look-up process similar to the operation of obtaining aerodynamic coefficients. The second type of operation is an integer SIN-COS table look-up such as might be utilized in the real-time simulation to gain a speed advantage. The third operation is a typical floating point function calculation typical of many similar operations in the real-time simulation. No attempt was made to optimize the FORTRAN benchmark as only the relative execution time is important. The computer manufacturers were given the benchmark and asked to run it on their respective machines "as is". The resultant execution time will provide one criterion to be used in computer choice. In addition, the manufacturers were invited to optimize the benchmark in FORTRAN, possibly using an optimizing compiler and return the optimized program and its execution time. Each manufacturer was also invited to write portions of the benchmark in assembler language so that the greater portion of the machines' power could be realized.

The basic benchmark will provide a common denominator for comparison of the machines' fundamental hardware/software combination. The optimized FORTRAN benchmark will provide an indication of the higher level software/hardware power of the machine. This is an important combination as it is desirable to write as much as is possible of the real-time simulation program in a higher level language. The assembler language benchmark would indicate machine power and programmer skill, but must be considered carefully so that the general similarity between the benchmark and the real-time simulation is not lost.

The results of the benchmark are not available at the time of this report. A listing of the benchmark is given in the Appendix.

As is well known, the actual computation portion of a computer system may comprise less than half of the total hardware system cost. If computer peripherals are expensive, they are also absolutely necessary for efficient utilization of the computer system. By not including critical software development peripherals in the computer system, one can trade a savings in hardware cost for many times this cost in additional software development expenses and delays. This same consideration holds for operating systems as they relate to software development.

The Anechoic Chamber facility currently maintains a substantial computer system in its equipment inventory. The efficient utilization of this equipment in the real-time simulation facility can play an important role in the cost-effectiveness of the real-time simulation facility. Investigation has led to the conclusion that the cost of interfacing the two computer systems would be excessive. This had led to the current attempts to transfer the tasks of the original computer system to the real-time simulation computer and recover the value of this original computer by trading it to a Wright Aeronautical Laboratory division user who can utilize it effectively.

This approach will lead to a substantial savings for the new owner of the existing Anechoic Chamber computer. It will lead to a substantial savings in interfacing cost for the Anechoic Chamber facility and will significantly reduce delay and software development cost for the real-time simulation facility. In addition, significant improvements in the present data acquisition task can be realized by transferring it to the new computer.

In summary the real-time simulation mini-computer must, above all, be extremely fast. In addition to this primary requirement, the computer system must be amenable to efficient software development and must be adaptable to secondary tasks.

SIMULATION SOFTWARE APPROACH

The software required of the real-time simulation is basically similar to the many analytical simulation packages indicated in the literature. The two important differences are: real hardware replaces certain software functions of these analytical simulations and the real-time program must be capable of updating all information within a certain time interval.

The time restriction means that the real-time software must be very time effective and in some cases may need to trade precision for time.

The presence of hardware-in-the-loop will tend to counter this trade-off as it neither requires nor produces extreme precision, and the presence of feed-back negates cumulative error effects.

Because of this similarity to existing analytical programs, a very cost-effective approach will be that of adapting useable modules of these programs. Essentially those modules which deal with the geometry, kinematics and dynamics, and target trajectory are adaptable. These modules represent a substantial portion of the real-time simulation software.

The functions of the real-time simulation software may now be given:

(1) Initialize the simulation
(2) Provide inertial reference inputs to the missile hardware
(3) Retrieve control surface commands from the missile hardware
(4) Update missile dynamics and missile/target geometry
(5) Produce relative missile-to-target position
(6) Provide target simulator (TAPS) with line-of-sight angle, range, and rate-of-change of range
(7) Record data which can be used to judge missile response
(8) Terminate simulation at the appropriate point prior to missile impact
(9) Compute miss distance by extrapolating from simulation parameters at termination

If the missile autopilot hardware is not available, this will have to be simulated by software.

System initialization refers to that effort required to produce the initial dynamic and geometric relationship between the target and missile and inertial reference. The missile hardware must be initialized to the appropriate state that it would acquire prior to playing an active role in guidance. The dynamic and kinematic characteristic of the missile must be passed to the real-time portion of the software so that inertial reference inputs to the missile hardware may be provided. The geometric relationship between missile and target must be passed to the target simulation so that the missile receives the desired signal upon initiating guidance. The missile hardware is then commanded to begin guidance and functions two through seven are executed during each update interval.

In providing inertial reference signals and retrieving control surface commands, the simulation software must pass appropriate parameters to IO modules. These IO modules will be written by the staff of the Anechoic Chamber to interface with the particular missile hardware. The update interval for this information has been chosen as 5 milliseconds. This time interval appears to be attainable by the simulation computer and appears to be small enough to satisfy missile dynamic requirements. Experimentation with this time interval will be possible after completion of the simulation system.

Computation of missile dynamics refers to computation of all information necessary to produce the missile attitude and position at each update time as a function of inertial forces, control surface movement and instantaneous momentum. Comparison of target attitude and position with missile attitude and position will produce the missile-to-target relative position.

When the line-of-sight range between the missile and target becomes less than some predetermined value or when the missile attitude to line-of-sight angle exceeds some predetermined value, the simulation will be terminated. The miss distance from this point will then be extrapolated from system parameters.

The first task in developing the real-time simulation software would be to produce computer code capable of accomplishing steps two through eight as discussed above within 64K bytes of computer main memory. This would be done without regard to the 5 millisecond update period. To assist in debugging a software simulator for the real-time simulation hardware and IO modules would be written. This simulator would be simplistic in nature and would not represent a specific missile. Its only purpose would be as a debugging aid.

To ensure a manageable task, the software should be written with a specific missile in mind. It is desirable to make the software generic in nature, but this should not be a specific objective of the initial software. After success of the initial simulation system has been proven, extension and modification of the software to make it generic would be a realistic objective.

A brief documentation effort would be included with this initial software. When the staff of the Anechoic Chamber receives this software, judgment on the cost and approach toward reaching a 5 millisecond update time, if not already attained, will be possible. Then additional effort and/or documentation can be arranged.

SUMMARY

The general nature of the real-time hardware-in-the-loop missile simulation has been discussed. The general equipment arrangement has been presented. Discussion of the equipment as it relates to requirements of the simulation computer was presented.

The criteria for choice of a computational approach has indicated that a digital mini-computer should be used. A detailed study of such computers was made and is summarized in this report. It was decided that the computer must be fast, easily maintainable, and adaptable to secondary existing and future tasks.

The general approach toward obtaining the simulation software has been discussed. It will be efficient to utilize portions of existing analytical simulation programs. The software must be adapted to run a relatively small computer and then must be optimized to execute in a minimum time. The general requirements of this software were discussed.

REFERENCES

1. Ball, R. F., A. W. Lee, and C. L. Lewis, "An Engineering and Programming Guide for a Six Degree of Freedom, Terminal Homing Simulation Program," TR-RG-73-22, MICOM Redstone Arsenal, Alabama, October 1973.

2. Data General Corporation - Programmer's Reference Manual Eclipse Line Computers. No. 015-000024, March 1975.

3. Digital Equipment Corporation - PDP 11/45 Processor Handbook, 1973.

4. First Ann Arbor Corporation, "Terminal Guidance Simulation User's Manual Model Report." Contract No. F29601-72-C-0156 Guidance Test Division, Holloman Air Force Base, June 1973.

5. Gill, P. E. "Description and Operation of the Generic Missile Model (GEMM) Simulation." Penval Working Paper No. 125, Calspan Corporation, Buffalo, New York. Contract F33615-73-C-4112, Air Force Avionics Laboratory, April 1975.

6. Goedeke, R. C. and C. W. Smoots, "AFAL Flight Simulator for ECM Evaluations." Report No. E6237-SR-2 Written by IIT Research Institute for Air Force Avionics Laboratory, September 1973.

7. Holmes, Wil ard, Charles M. W ll, and Alexander C. Jolly, "Time Critical Hybrid Computer Simulation for Missile Systems with Hardware-in-the-Loop Operation." TR-RG-73-1, January 1973.

8. Lange, A. S., "Semi-Physical Simulation of Guided Missiles." Computers and Elect. Engng," Vol. 1, pps. 119.142, Pergam n Press, 1973, printed in Great Britain.

9. Modular Computer Systems, Inc. Reference Manual Modcomp IV Central Processor. No. 210-110000-000.

10. Reese, Stan, "Adaptable Mini-computer-Based Test Systems." Data Systems Engineering. Anaheim, California.

11. Rogers, Gene B., "Memo: Boeing Terminal Guidance Laboratory Simulation Experience, Field-of-View Requirement." Boeing, July 7, 1975.

12. Varian Data Machines. Varian 74 System Handbook. Irvine, California, February 1975.

APPENDIX

A computer listing of the benchmark program is contained in this Appendix.

```
      COMMON ITTYI,ITTYO,ILPT
      COMMON IABLE(1024),FINC
      COMMON A1(10,10),A2(10,10),A3(10,10),X1(10),Y1(10)
      COMMON X2(10),Y2(10),X3(10),Y3(10),CUAL(1000)
      F1(X,Y)=6.38*X+10.5*Y+500.
      F2(X,Y)=3.10*X-4.00*Y+500.
      F3(X,Y)=16.0*X+23.0*Y+500.
C
C
C     THIS IS THE MAIN BENCHMARK PROGRAM. IT CONTAINS AN
C     INITIALIZATION SECTION, A CALL TO THE TIMED CALCULATIONS,
C     AND AN OUTPUT SECTION. NO TIMING DATA IS DESIRED FOR ANY
C     OTHER THAN THE CALCULATION SECTION. TIMING START AND
C     STOP POINTS ARE INDICATED BY COMMENT STATEMENTS. IF A
C     FORTRAN CALLABLE TIMING ROUTINE IS AVAILABLE, IT MAY BE
C     INSERTED AT THE APROPRIATE POINTS. NO OTHER MODIFICATIONS
C     SHOULD BE MADE TO THIS SECTION OF THE PROGRAM.
C
C
C
C
C     THE VARIABLES ITTYI,ITTYO,ILPT ARE LOGICAL UNIT NUMBERS OF
C     TELETYPE INPUT, TELETYPE OUTPUT, AND LINE PRINTER OUTPUT
C     RESPECTIVELY. ALL LINE PRINTER OUTPUTS ARE TTY COMPATIBLE
C
C
      ITTYI=1
      ITTYO=2
      ILPT=2
      FINC=2.0**(-10)
      R1=FINC/16.
      RAD=0.
      DO 200 J=1,1024
      X=FLOAT(J)*FINC
100   RAD=RAD+R1
      Y=SIN(RAD)
      IF(X-Y)200,200,100
200   IABLE(J)=RAD*57.29577*60.
300   FORMAT(10F7.3)
      F=.33333
      X1(1)=F
      X2(1)=F
      X3(1)=F
      F=1.0
      Y1(1)=F
      Y2(1)=F
      Y3(1)=F
      X1(2)=141.
      X2(2)=302.
      X3(2)=261.
      Y1(2)=9.
      Y2(2)=2.
      Y3(2)=5.
      X1(3)=505.
      X2(3)=543.
      X3(3)=299.
      Y1(3)=14.
      Y2(3)=3.
      Y3(3)=11.
      X1(4)=818.
      X2(4)=732.
      X3(4)=361.
      Y1(4)=17.
      Y2(4)=7.
      Y3(4)=22.
      X1(5)=1155.
      X2(5)=854.
      X3(5)=462.
      Y1(5)=24.
      Y2(5)=18.
      Y3(5)=34.
      X1(6)=1159.
      X2(6)=1314.
      X3(6)=523.
      Y1(6)=27.
      Y2(6)=25.
      Y3(6)=37.
      X1(7)=1522.
      X2(7)=1694.
      X3(7)=722.
      Y1(7)=44.
      Y2(7)=41.
      Y3(7)=38.
      X1(8)=1660.
      X2(8)=1960.
      X3(8)=813.
      Y1(8)=48.
      Y2(8)=42.
      Y3(8)=55.
      X1(9)=2052.
      X2(9)=2151.
      X3(9)=1471.
      Y1(9)=59.
      Y2(9)=65.
      Y3(9)=63.
      F=3335.
      X1(10)=F
      X2(10)=F
      X3(10)=F
      F=106.
      Y1(10)=F
      Y2(10)=F
      Y3(10)=F
      DO 900 J=1,10
      DO 700 I=1,10
700   A1(I,J)=F1(X1(I),Y1(J))
      DO 800 I=1,10
800   A2(I,J)=F2(X2(I),Y2(J))
      DO 900 I=1,10
900   A3(I,J)=F3(X3(I),Y3(J))
      WRITE(ILPT,1000)((A1(I,J),I=1,10),J=1,10)
      WRITE(ILPT,1000)((A2(I,J),I=1,10),J=1,10)
      WRITE(ILPT,1000)((A3(I,J),I=1,10),J=1,10)
1000  FORMAT(10(10F7.0,/),//)
      WRITE(ILPT,1050)((X1(I),I=1,10),(Y1(I),I=1,10))
      WRITE(ILPT,1050)((X2(I),I=1,10),(Y2(I),I=1,10))
      WRITE(ILPT,1050)((X3(I),I=1,10),(Y3(I),I=1,10))
1050  FORMAT(2(10F7.0,/),//)
C
```

```
C
C       THE BENCHMARK TIMING SHOULD START AT THIS TIME
C
C
        WRITE(ITTYO,11ØØ)
11ØØ    FORMAT("SETUP COMPLETE: START TIMER"/)
        CALL DATA
C
C
C       THE BENCHMARK TIMING SHOULD STOP AT THIS POINT
C
C
        WRITE(ILPT,3ØØ)(CUAL(L),L=1,1ØØØ)
        END
        END$

-OF-TAPE
```

```
      SUBROUTINE DATA
      COMMON ITTYI,ITTYO,ILPT
      COMMON IABLE(1024),FINC
      COMMON A1(10,10),A2(10,10),A3(10,10),X1(10),Y1(10)
      COMMON X2(10),Y2(10),X3(10),Y3(10),CUAL(1000)
C
C
C     THESE ARE THE BENCHMARK TIMING ROUTINES. THEY CAN BE
C     MODIFIED SO LONG AS THEY REMAIN FUNCTIONALLY INTACT.
C     ASSEMBLY LANGUAGE VERSIONS CAN BE SUBSTITUTED WHERE DESIRED
C
C
      DELQ=216.3184
      DELR=28.4963
      CPREV=0
      CPRE1=0
      DO 150 K=1,10
      DO 150 I=1,100
      DO 150 J=1,100
      X=FLOAT(I)*FLOAT(J)*.33333
      Y=FLOAT(I+J)/FLOAT(J)
      C=FUNCT(A1,X1,Y1,X,Y)+(CPREV-CPRE1)*TSIN(Y)**2+FUNCT(A2,X2Y
     1Y)*DELQ*TCOS(Y)-FUNCT(A3,X3,Y3,X,Y)*DELR*TSIN(X)
      CPREV=CPRE1
      CPRE1=C
      M=I*J
      IF(M-1000)26,26,150
26    CUAL(M)=C
150   CONTINUE
      RETURN
      END
      FUNCTION TOUT(IX)
      COMMON ITTYI,ITTYO,ILPT
      COMMON IABLE(1024),FINC
      LSIZE=512
      LPOIT=512
10    LSIZE=LSIZE/2
      IF(IABLE(LPOIT)-IX) 50,30,30
30    LPOIT=LPOIT-LSIZE
      GO TO 100
50    LPOIT=LPOIT+LSIZE
100   IF(LSIZE-1)10,65,10
65    TOUT=FLOAT(LPOIT)*FINC
      RETURN
      END
      FUNCTION TCOS(X)
      TCOS=TSIN(90.-X)
      RETURN
      END
      FUNCTION TSIN(X)
      Y=ABS(X)
      IX=AMOD(Y,360.)*60.
      I=1+IX/5401
      IF(X)200,9,9
9     GO TO (10,20,30,40),I
10    TSIN=TOUT(IX)
      RETURN
20    TSIN=TOUT(10800-IX)
      RETURN
```

```
30    TSIN=-TOUT(IX-10800)
      RETURN
40    TSIN=-TOUT(21600-IX)
      RETURN
200   GO TO (110,120,130,140),I
110   TSIN=-TOUT(IX)
      RETURN
120   TSIN=-TOUT(10800-IX)
      RETURN
130   TSIN=TOUT(IX-10800)
      RETURN
140   TSIN=TOUT(21600-IX)
      RETURN
      END
      FUNCTION FUNCT(FJ,X1,X2,X,Y)
      DIMENSION FJ(10,10),X1(10),X2(10)
      COMMON ITTYI,ITTYO,ILPT
      DO 20 I=2,10
      IF(X1(I)-X)20,19,19
19    I1=I
      GO TO 25
20    CONTINUE
      GO TO 1000
25    DO 30 I=2,10
      IF(X2(I)-Y)30,29,29
29    I2=I
      GO TO 35
30    CONTINUE
      GO TO 1000
35    IX=I1-1
      IY=I2-1
      F=FJ(I1,I2)
      FUNCT=F+(FJ(IX,I2)-F)/(X1(IX)-X1(I1))*(X-X1(I1))+
     1(FJ(I1,IY)-F)/(X2(IY)-X2(I2))*(Y-X2(I2))
      RETURN
C
C
C     THIS ERROR INDICATION NEED NOT BE OPTIMIZED, BUT SHOULD BE
C     INCLUDED.
C
C
1000  WRITE(ITTYO,1001)
1001  FORMAT("ERROR IN FUNCT"//)
      STOP
      END
      END$

-OF-TAPE
```

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT-PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

FATIGUE CRACK PROPAGATION IN LAMINATED AND MONOLITHIC ALUMINUM ALLOY PANELS

Prepared by: J. A. Alic

Academic Rank: Assistant Professor

Department & University: Department of Mechanical Engineering
Wichita State University

Assignment:
(Laboratory) - Flight Dynamics
(Division) - Structures
(Branch) - Structural Integrity

USAF Research Colleague: J. P. Gallagher

Date: 15 August 1975

Contract no: F44620-75-C-0031

FATIGUE CRACK PROPAGATION IN LAMINATED AND MONOLITHIC ALUMINUM ALLOY PANELS

by
J. A. Alic

ABSTRACT

Adhesive bonding is increasingly being used for joining structural components of aircraft. As applications spread to primary structure, it becomes important to be able to characterize rates of fatigue crack growth in bonded components.

For this study, two particular well-defined crack geometries -- panels with either through-the-thickness center cracks or radial cracks emanating from holes -- were used to compare fatigue crack growth rates in laminated plates with those in monolithic plates. The material was 7075-T651 aluminum alloy; laminated panels were bonded with AF-55 adhesive and had either two or four layers, the overall thickness, 1/2 in., being the same as for the monolithic specimens. Testing was at constant load amplitude.

Crack propagation data show that the growth behavior in the laminated panels is not, in general, significantly different from that exhibited by monolithic material. These results reflect the relatively large thicknesses of the laminae used -- the thinnest being 1/8 in. -- and the fact that the variation in thickness was a factor of only four. An important conclusion of the study is that fatigue crack growth data from monolithic material can be used for design purposes to analyze adhesively bonded structure in the presence of through-the-thickness cracks or flaws.

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

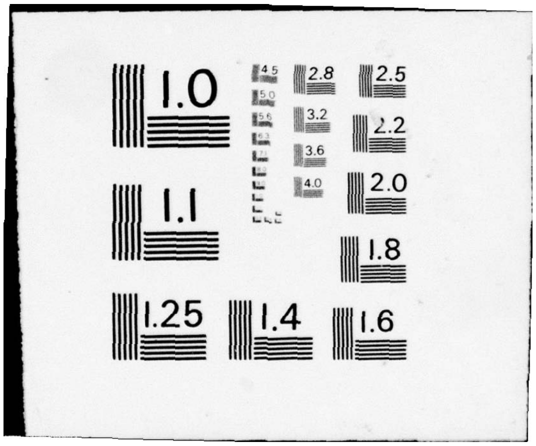
Although considerable experience in aircraft applications has been accumulated with the use of adhesive bonding for joining components of secondary structure -- such as control surfaces -- applications to primary load-carrying structure are in the development stage, particularly for military aircraft. As a prerequisite to adoption of bonding for safety-of-flight structure, knowledge of fatigue crack growth behavior in bonded structure is needed. While many different combinations of crack and structural geometry will have to be considered prior to design analyses, the purpose of the present study was to generate baseline crack growth data for adhesively bonded elements. To this end, a pair of simple geometries have been used, both of which involve cracks extending fully through the thickness of bonded panels.

A major impetus behind development of adhesively bonded structures for aircraft stems from prospects for decreased cost and weight through the elimination of mechanical fasteners such as bolts and rivets. Dispensing with mechanical fasteners may also help to provide longer life structures because of the propensity of fatigue cracks for starting at fastener holes. Use of adhesive bonding for primary structure of long-lived aircraft will require knowledge of the strength and fracture resistance of the bonded joint itself, particularly in the presence of various potentially degrading environments; however the present effort is concerned with the crack growth behavior in the metallic elements joined by the adhesive not with the properties of the adhesive itself.

In addition to possible cost and weight savings, and the elimination or reduction in the number of fastener holes, use of bonded structure may also serve to increase damage tolerance. This is due to the influence of weakly-bonded interfaces which can act as crack-dividers or crack-arresters, or both, depending upon the orientation of the bonded interface with respect to the crack front. In the crack-divider configuration, the crack front is perpendicular to the bond-line, while in the arrester case, it is parallel to the bond-line. Thus a crack-divider geometry gives rise to two or more cracks moving somewhat independently of one another because separated by the joint, which has, typically, low tensile strength compared to the metallic laminae. In the crack-arrester case, the progress of the crack will be directly impeded by the lack of continuity at the interface. These, of course, are extremes, and naturally occurring cracks in bonded structures will often have mixed character. However both configurations are likely to have some efficacy in retarding both subcritical crack growth, from whatever cause, and final failure in the presence of damage. To take advantage of such behavior it is necessary only that the interface be weak compared to the materials being joined; diffusion bonding, brazing, roll bonding and other methods may be used as well as adhesive bonding.

AD-A031 017   AUBURN UNIV ALA   SCHOOL OF ENGINEERING   F/G 5/2
1975 USAF/ASEE SUMMER FACULTY RESEARCH PROGRAM.(U)
SEP 75   J F O'BRIEN   F44620-75-C-0031
UNCLASSIFIED   AFOSR-TR-76-1113   NL

5 OF 7
ADA031017


1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

In fact, weak layers in monolithic material such as might be caused by segregated impurities can have the same effect. A brief review of past work concerning the influence of lamination on both fatigue and fracture toughness is included in Ref. 1. The present study has been concerned with fatigue crack growth in a pure crack-divider orientation.

TEST SPECIMENS

The material chosen for this program was 7075-T651 aluminum alloy, widely used in aircraft structures. All test specimens were fabricated from a single half-inch thick plate. Mechanical properties measured on a pair of tensile specimens cut from this plate are given in Table 1. While the simple tensile properties are normal for this material, the tensile bars, which were cut so that the loading was in the rolling direction, as for the fatigue specimens, showed considerable splitting and delamination upon fracture. The splits, which were parallel to the plate surface, indicate that the plate was heavily worked in the last pass during rolling. It is also possible that stringering of included particles contributed to this behavior.

Table 1

Mechanical Properties of 7075-T6
Aluminum Used for Fatigue Test Panels

Yield Strength	75 ksi
Tensile Strength	82 ksi
Elongation at Fracture in 2 in.	9%
Fracture Toughness	45 ksi $\sqrt{in}$ *

All values averages from two specimens.

* Approximate; determined by loading two center-cracked monolithic (1/2 in. thick) fatigue specimens to failure following fatigue testing (see also Table 5).

A total of twelve fatigue test specimens were fabricated from the 1/2 in. plate, two each of six configurations, as outlined in Table 2; while time constraints have made it impossible to complete the testing of all twelve specimens as of the date of this report, all twelve are listed for completeness. Of the twelve panels, six contained a pair of central slots, Fig. 1, while the other six had three axially separated holes, Fig. 2, from which radial slots were cut as crack-starters. The center-cracked specimens are representative of a configuration widely used for the generation of basic crack propagation data, while the cracked-hole panels simulate the case of a crack leading from a fastener hole, a more realistic situation from the standpoint of actual aircraft structures.

Bonding of the laminated panels was performed by a vendor* after machining of the laminae. The adhesive was AF-55**, a supported film epoxy. Surface preparation and bonding procedure followed standard aircraft practice. An FPL etch was used; curing was accomplished by vacuum bagging the laminates and holding at 250°F and 40 psi in an autoclave for one hour. The bondline thicknesses averaged 0.004 in.

TEST PROCEDURE

All testing was performed using sinusoidal loading in a servo-controlled electro-hydraulic testing machine at a cyclic frequency of 5 Hz. Panels were loaded in fluctuating tension with a load ratio $R=P_{min}/P_{max}$ of 0.1. The tests were carried out in laboratory air of generally high humidity. Cracks were initiated from the spark-machined or saw-cut slots using the same loads as for the crack growth testing. The test loads and the corresponding nominal stresses are listed in Table 3. Only those panels for which testing was complete as of the date of this report are included in the table.

Crack lengths were measured by means of mylar tapes calibrated in 0.005 in. increments attached to each side of the specimen parallel to the crack paths. The crack lengths were read from the tapes with low-power microscopes. In all cases, crack lengths on both sides of

* Materials Research Laboratory, Inc., Glenwood, Illinois.

** 3M Company.

Table 2. Test Panels.

Specimen Type	Number of Panels	Number of Layers	Nominal Layer Thickness (in.)	Crack Configuration	Description
Monolithic	2	1	1/2	center-crack	Fig. 1
Monolithic	2	1	1/2	cracked-hole	Fig. 2
Laminated	2	2	1/4	center-crack	Fig. 1
Laminated	2	2	1/4	cracked-hole	Fig. 2
Laminated	2	4	1/8	center-crack	Fig. 1
Laminated	2	4	1/8	cracked-hole	Fig. 2

Total thickness of each specimen is 1/2 in.

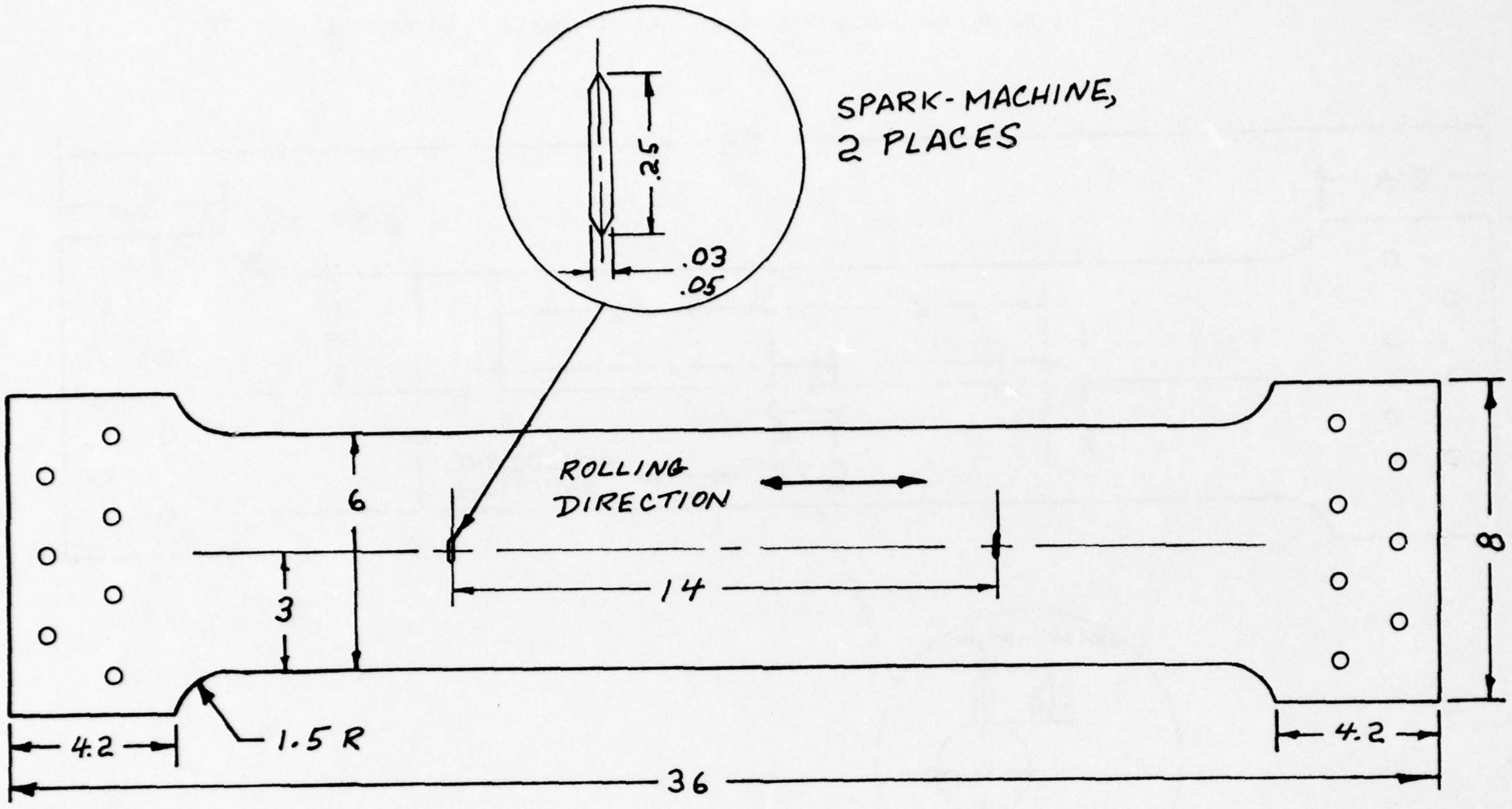


Fig. 1. Center-Cracked Test Panel.

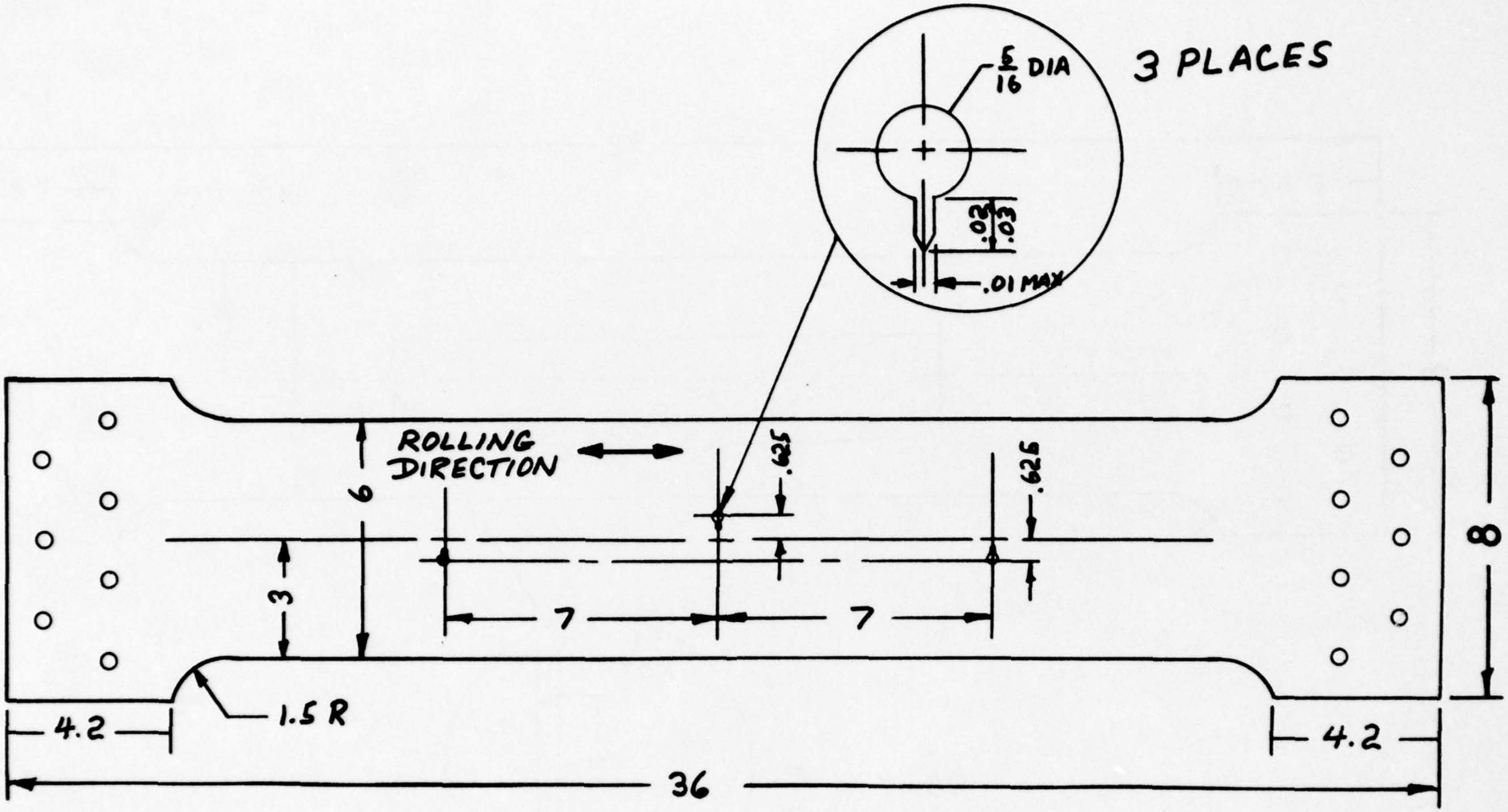


Fig. 2. Test Panel with Radial Crack Emanating from Circular Hole.

Table 3. Test Conditions.

Specimen Type	Specimen Designation	Number of Fatigue Cracks	Load (Kips)		Nominal Stress (ksi)	
			Max	Min	Max	Min
Monolithic, Center-cracked	MCC-1	2	20	2.0	6.7	0.67
Monolithic, Center-cracked	MCC-2	2	24	2.4	8.1	0.81
Monolithic, Cracked-hole	MHC-1	3	30	3.0	10.0	1.0
Monolithic, Cracked-hole	MHC-2	3	24	2.4	8.1	0.81
Two-Layer Laminate, Center-cracked	LCC-2-1	2	24	2.4	8.0	0.8
Two-Layer Laminate, Center-cracked	LCC-2-2	2	24	2.4	8.0	0.8
Four-Layer Laminate, Center-cracked	LCC-4-1	2	24	2.4	8.0	0.8
Four-Layer Laminate, Center-cracked	LCC-4-2	2	24	2.4	8.0	0.8

the panels were measured; the results for each crack location could then be analyzed by considering panel sides individually as well as by averging the measurements from both sides. The crack length measurements are estimated to be accurate to within ± 0.002 in., while the specimen loads could be controlled to within ∓ 1/2%.

RESULTS AND DISCUSSION

The fatigue crack propagation data were analyzed by correlating growth rate da/dN, where a is crack length (cracked-hole panels) or half-crack length (center-cracked panels) and N number of fatigue cycles, with the maximum value of the stress intensity factor, K_{max}, during cycling. Here

$$K_{max} = \sigma_{max} \sqrt{\pi a} \cdot \beta \qquad (1)$$

In this equation σ_{max} is the greatest nominal stress during a cycle, constant throughout a test, and β is a factor which depends upon panel geometry.

The data reduction was performed using a computer program* which calculates da/dN for each crack length by fitting a least-squares linear equation to the seven data points of which that particular crack length is the average calculated crack length. The slope of the line is taken as da/dN for that crack length. The corresponding K_{max} value is calculated from the applied loads, specimen dimensions and crack length. For the center-cracked panels a secant finite-width correction factor (Ref. 2) is used, so that

$$K_{max} = \sigma_{max} \sqrt{\pi \sec (\pi a/W)} \qquad (2)$$

W being the panel width. The cracked-hole case is treated by means of the equation

$$K_{max} = \sigma_{max} \sqrt{\pi a} [0.6762 + 0.8733/(0.3245 + a/r)] \qquad (3)$$

where r is the radius of the hole. The term in brackets corresponding to the β of Eq. 1 is the result of a curve fit to the stress intensity solution of Bowie (Ref. 3). Among the results of the computer analysis is then a listing of the values of da/dN and K_{max} which correspond to the crack length readings. The program can also be used to plot these results.

In addition, the analysis program fits a least-squares straight line to the da/dN-K_{max} data corresponding to the power law expression

$$da/dN = C\, K_{max}^{n} \qquad (4)$$

where C and n are experimentally determined constants. Relations

* Developed by J. P. Gallagher, Fatigue, Fracture and Reliability Group, Air Force Flight Dynamics Laboratory.

corresponding to Eq. 4 are then available for each test panel -- for each side individually and for both sides together. Because the variations from one side of a panel to the other were in general small, only the average results from the measurements on both sides are discussed below. The crack growth rate data reduction program can also be used to treat groups of data from a number of test specimens; in this way it is possible, for example, to compare all monolithic panels to various groupings of laminated panels.

Monolithic Specimens

The four monolithic panels of full 1/2 in. thickness were intended to provide a baseline to which the laminated specimens could be compared. The crack growth rate data for these four specimens are shown in Fig. 3. This figure includes measurements from a total of ten cracks -- two each in the two center-cracked panels, three from each of two cracked-hole panels -- at three different stress levels (Table 3). The results fall within the expected range for this material (Ref. 4). There are no discernible systematic differences between the center-cracked and the cracked-hole specimens. The least-squares power law fit to all the monclithic panel data shown in Fig. 3 is included in Table 4.

Laminated Specimens

Four laminated panels have been tested -- one pair having two 1/4 in. thick layers and another pair made with four 1/8 in. thick layers. All four were center-cracked (Fig. 1) and all were cycled at the same stress levels -- 0.8 ksi to 8.0 ksi (Table 3). Thus data from four cracks is available for 2-layer laminates and also for 4-layer laminates.

The 2-layer laminate results are shown in Fig. 4. As for the monolithic panels, the crack lengths measured on both sides of the panels have been averaged. The power law equation corresponding to the results shown in Fig. 4 is given in Table 4.

Table 4. Least-Squares Fits to Fatigue Crack Propagation Data. (da/dN in in./cycle, K_{max} in ksi$\sqrt{in.}$)

Specimen Group	Equation
All monolithic (2 center-cracked, 2 hole-cracked)	$da/dN = (1.69 \times 10^{-9})\, K_{max}^{3.72}$
2-layer laminates (2, center-cracked)	$da/dN = (1.91 \times 10^{-10})\, K_{max}^{4.54}$
4-layer laminates (2, center-cracked)	$da/dN = (2.68 \times 10^{-10})\, K_{max}^{4.41}$

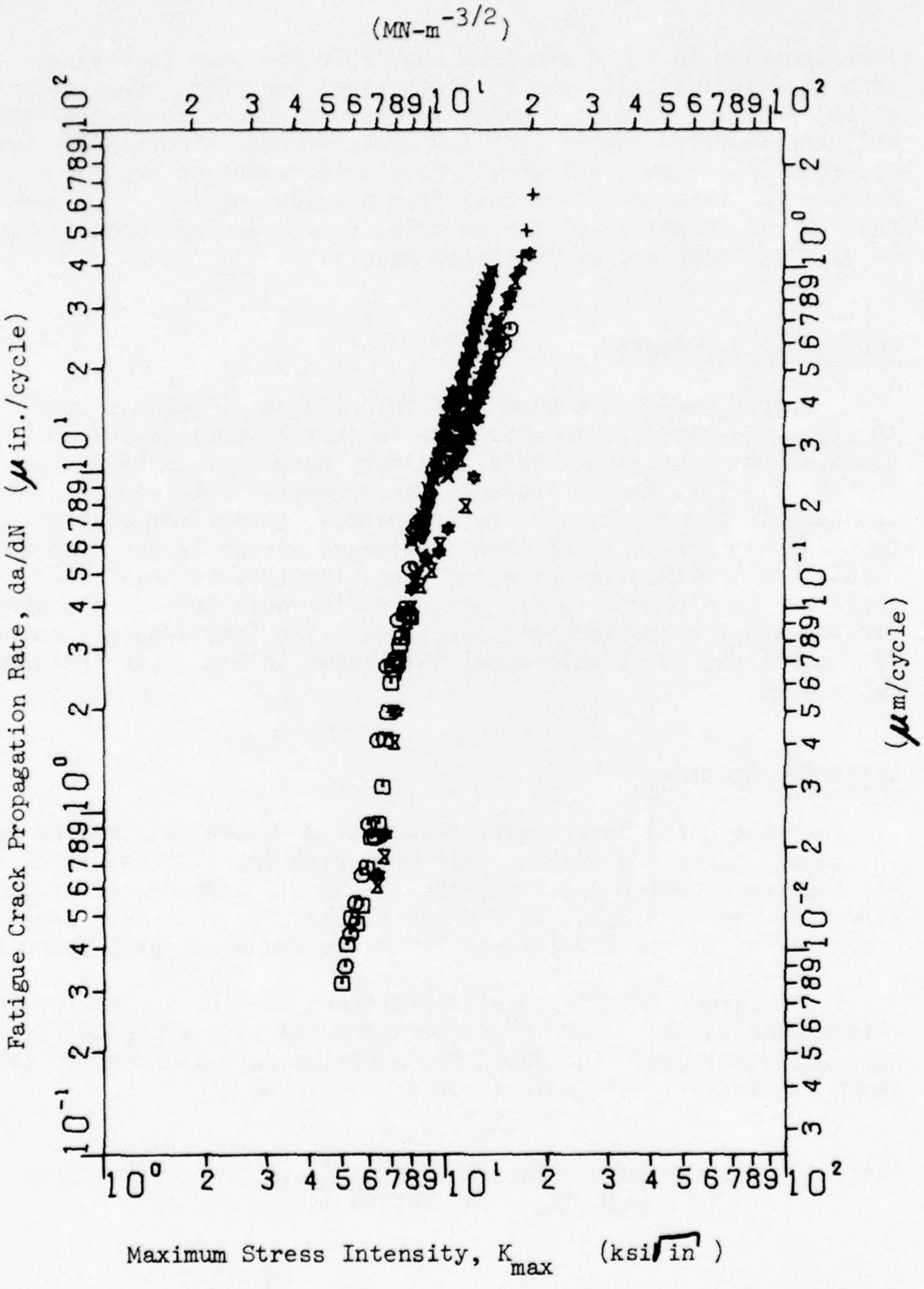


Fig. 3. Fatigue Crack Propagation Rates in Monolithic Panels. Averages from Both Sides of Panels. (R = 0.1; Thickness 1/2 in.)

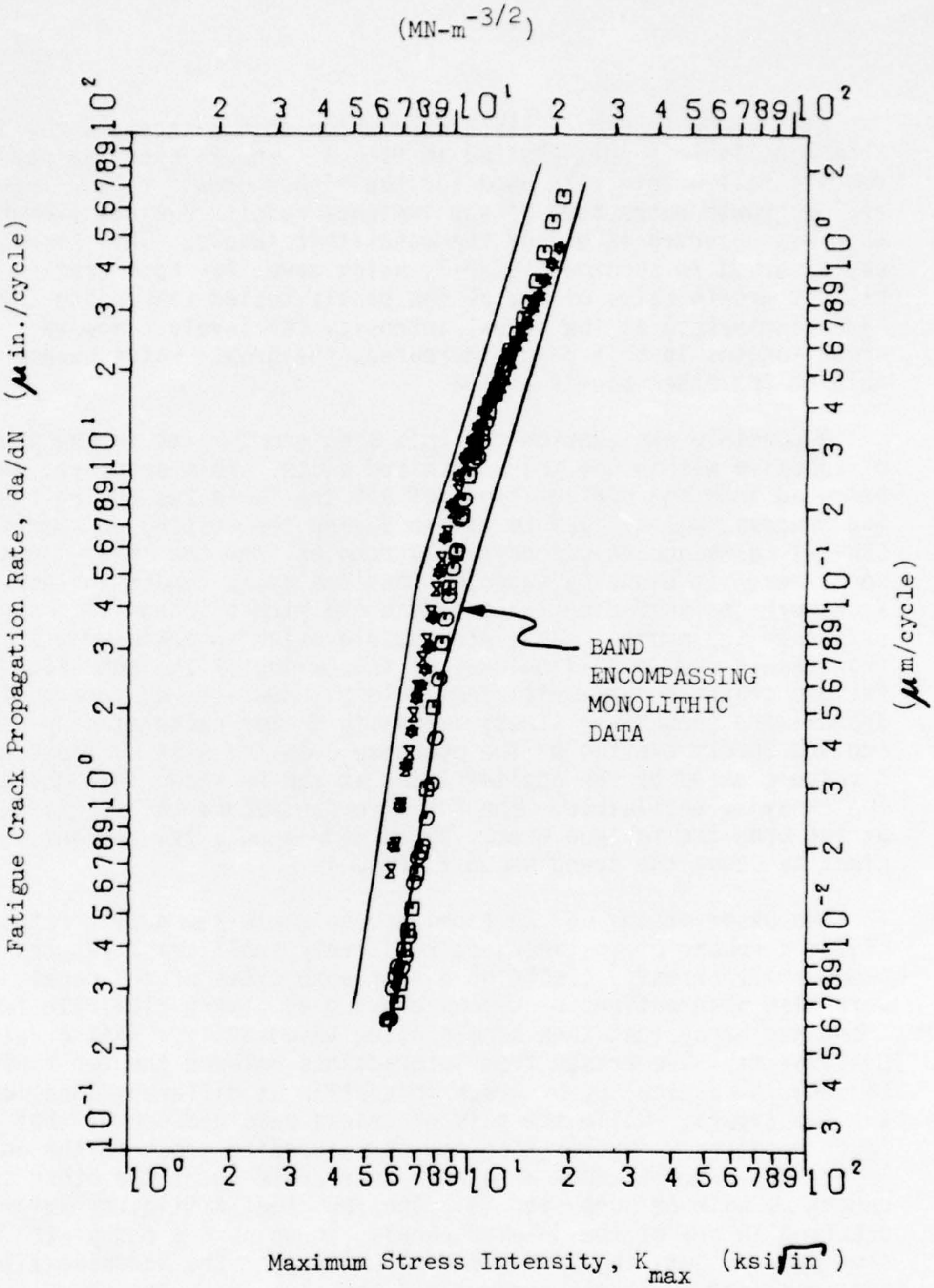


Fig. 4. Fatigue Crack Propagation Rates in 2-Layer Laminates. Averages from Measurements on Both Sides of Panels. (R = 0.1; Overall Thickness 1/2 in.)

Also shown in Fig. 4 is a band which encompasses all the data from monolithic panels plotted in Fig. 3. The 2-layer laminate results fall within this bond for the higher growth rates. However at low growth rates some of the laminate results exhibit slower growth than was observed in any of the monolithic panels. This slow growth was observed in specimen LCC-2-1, which gave, for both cracks, the slowest growth rates of any of the panels tested (including the 4-layer laminates) at low stress intensity (K) levels. However, as the crack lengths in this panel increased, the growth rates became comparable to the other panels tested.

A possible explanation for this slow growth lies in the presence of adhesive within the spark-machined slots. This adhesive, which was extruded into the pre-cut slots of all the laminates during the bonding process, was allowed to remain during the testing of specimen LCC-2-1, although it was partially removed from the other laminated specimens. It might be supposed that the epoxy tended to restrain the slot surfaces sufficiently to reduce the stress intensities serving to propagate the cracks. The later acceleration in crack growth could then result from a pulling away or fracturing of the adhesive when the fatigue cracks became sufficiently long. However, by superposing an approximate (negative) stress intensity factor calculated by estimating the forces exerted by the adhesive upon the slot surfaces upon the K value created by the applied load, it can be shown that the effect of the epoxy is negligible. The adhesive can reduce the stress intensities acting upon the fatigue cracks by no more than a few percent, insufficient to cause the trend shown in Fig. 4.

The observations on the panel giving these low growth rates included a number of instances at relatively small crack lengths of temporarily arrested cracks on one or both sides of the panel. There were also observations of cracks growing at a very slow rate for many thousands of cycles, then accelerating temporarily. This erratic behavior may have arisen from interactions between the two laminae, caused, in particular, by crack initiation at different locations in the two layers. While the pair of cracks remained short, that in one layer might grow further than the corresponding crack in the adjacent layer, for example, then arrest or decelerate until the other crack caught up with or surpassed it. The fact that similar behavior occurred in one of the 4-layer panels, in which the epoxy within the slot was cut out, supports the conclusion that the adhesive-filled slot does not by itself account for the slow crack growth.

One of the cracks in a 2-layer panel grew with a length consistently longer in one layer than in the other. The fracture surface, Fig. 5, showed that the composite crack front possessed a stepped character rather then running diagonally through the panel. There was also a small difference in level evident between the two layers on the

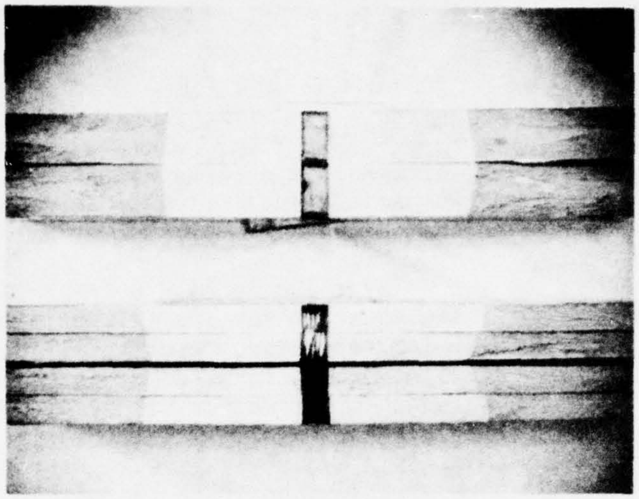


Fig. 5. Fracture Surfaces in Monolithic Panel and in 2-Layer Laminate.

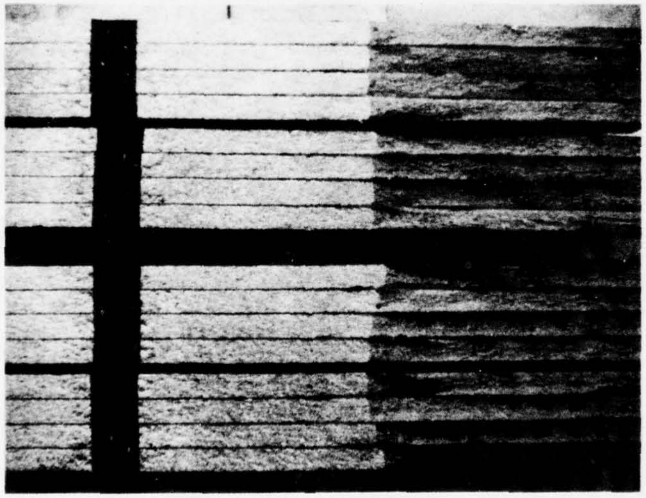


Fig. 7. Fracture Surfaces of 4-Layer Laminates.

fracture surface. Similar behavior observed in the 4-layer laminates shows that fatigue cracks in individual layers can grow with a considerable degree of independence.

Results from the two 4-layer laminates are presented in Fig. 6. The least-square power law equation for the 4-layer panels is included in Table 4. The 4-layer laminates gave results falling predominantly within the band of monolithic results, with, again, somewhat slower growth for small crack lengths. It should be noted, however, that there is less data available for monolithic material at low growth rates than for the laminates. Crack growth was, as for the 2-layer laminates, somewhat more erratic than in the monolithic material at short lengths. In particular, the 4-layer specimens exhibited instances of cracks growing faster on one side of the panel than on the other by a factor of more than two. This behavior probably reflects the influence of an uneven crack front through the four laminae, caused again by different crack origins in the several layers. As these cracks increased in length, the discrepancies from side to side decreased. For those cracks which showed such behavior, there were also frequent instances of temporary arrest and/or very slow growth.

The fracture surfaces of the 4-layer laminates, Fig. 7, show a definite stepped character; however, the variation from layer to layer is relatively small compared to the overall crack length. Thus the fatigue crack growth behavior shown in Fig. 6 can be considered representative even though crack lengths were measured only on the panel surfaces.

The changes in slope present at growth rates of about 10 μin/cycle in both sets of laminate data, Figs. 4 and 6, are also present in the monolithic data plotted in Fig. 3. Such slope changes are typically found in this growth rate region for 7075-T6(51) aluminum alloy (Refs. 4 and 5), although often at somewhat higher growth rates. While a slope decrease such as seen in Fig. 4 or Fig. 6 is sometimes associated with a transition from a flat fracture mode to a slant fracture mode (Ref. 5), shear lips only began to develop in the laminated panels at growth rates an order of magnitude higher than that of the slope change. No shear lip development was evident in the fatigue region of the monolithic panels.

The power law equations listed in Table 4 are similar for laminates of two and four layers. However the exponent n is smaller for the monolithic panels than for the laminates. This is caused by the preponderance of high crack growth rate monolithic data, much of which comes from the cracked hole specimens. This high growth rate data falls above the slope change region; as the data trend in Figs. 3, 4 and 6 does not conform with a power law description, these equations are presented only for comparison with other fatigue crack growth rate equations available in the literature.

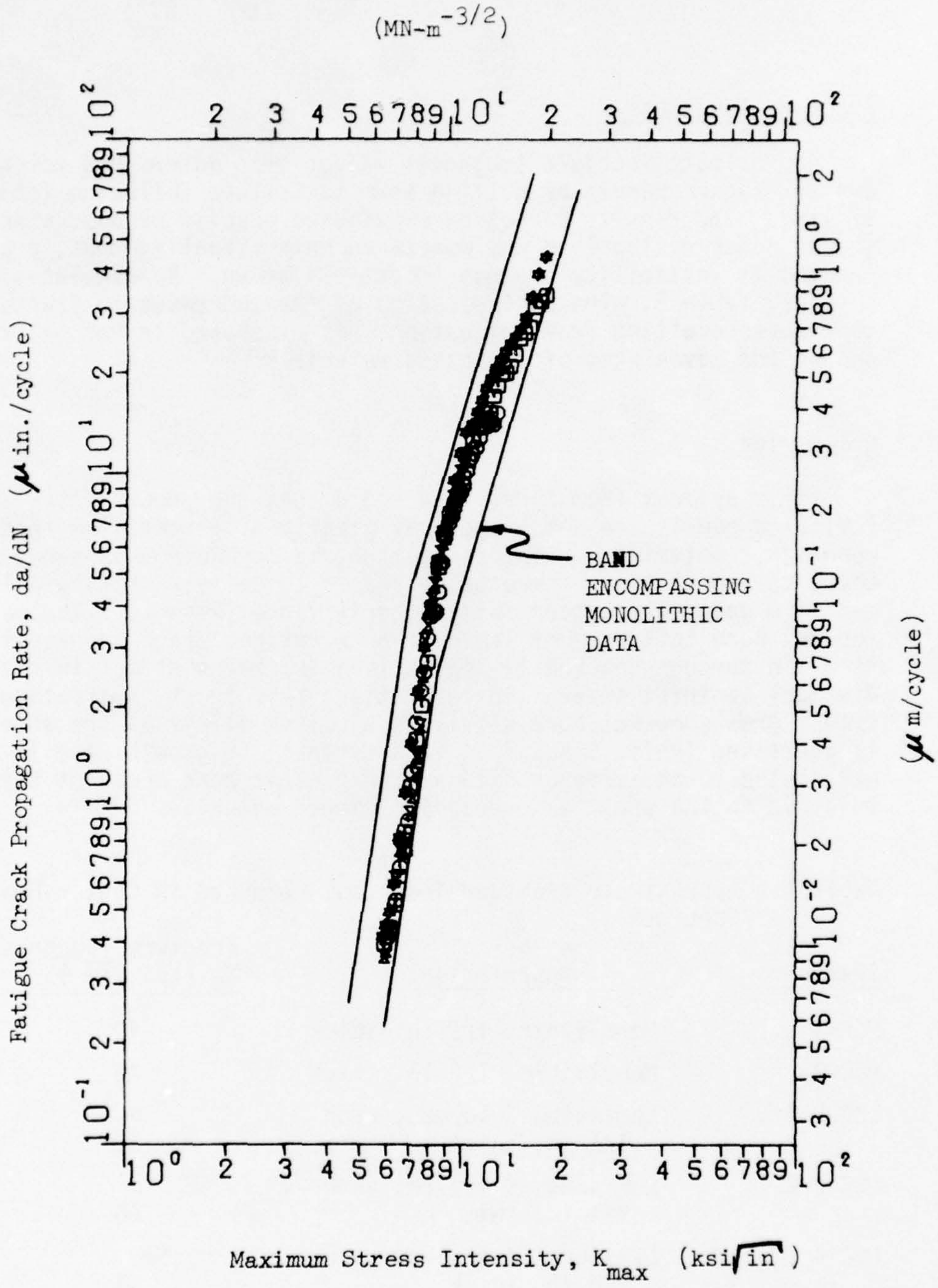


Fig. 6. Fatigue Crack Propagation Rates in 4-Layer Laminates. Averages from Measurements on Both Sides of Panels. (R = 0.1; Overall Thickness 1/2 in.)

Fracture Toughness

Approximate fracture toughness values were determined for all the center-cracked panels by pulling them to failure following fatigue cycling. The results cannot be considered precise because crack growth under rising load was monitored only visually; thus, crack lengths at instability are not accurately known. Nonetheless, the results, Table 5, give an indication of the increases in fracture toughness resulting from lamination. As discussed in Ref. 1, this is one of the advantages of laminated materials.

Discussion

It is evident from Figs. 3, 4 and 6 that the growth rates in the laminated panels are not in general greatly different from those in monolithic material. This result might be considered somewhat contrary to expectation, there being reason to believe that thin laminae can give generally better fatigue performance (Ref. 1). The basis for any such influence of lamination on fatigue crack propagation rates in through-cracked laminates is a thickness effect in the individual laminae; several investigations have shown that fatigue cracks grow somewhat more slowly in aluminum alloys as the sheet thickness is decreased (Refs. 6 and 7). This decrease in growth rate is usually attributed to the greater size of the plastic zone ahead of the crack relative to the sheet thickness in thinner material.

Table 5. Approximate Fracture Toughness Measured in Center-Cracked Specimens.

Specimen	Description	Fracture Toughness K_c (ksi $\sqrt{in}$)
MCC-1	Monolithic, 1/2 in. thick	43
MCC-2	Monolithic, 1/2 in. thick	46
LCC-2-1	Laminate, 2 layers, each 1/4 in. thick	50
LCC-2-2	Laminate, 2 layers, each 1/4 in. thick	60
LCC-4-1	Laminate, 4 layers, each 1/8 in. thick	61
LCC-4-2	Laminate, 4 layers, each 1/8 in. thick	60

Although the sources of thickness effects on fatigue crack propagation are somewhat obscure, it is noteworthy that the growth rate differences reported in Refs. 7 and 8 have been restricted to sheet material of thickness less than 0.120 in. Only for the 4-layer panels of the present study are the laminae in this range; it can be concluded that the thicknesses of the laminae used in the present study are greater than those for which pronounced effects on growth rate would occur and that the congruence in growth rates between monolithic and laminated panels is to be expected.

Implicit in the above discussion is the assumption that the lamina thickness rather than the overall thickness will control the crack propagation rate. For this to occur, the inter-layer bonding must be weak compared to the yield strength of the laminae so that relaxation and/or debonding can occur in the region ahead of the crack. This condition appears to have been met in the present case, the shear strength of the AF-55 epoxy being of the order of 5000 psi. The occasional differences in elevation between laminae observed on fatigue fracture surfaces, together with the stepped crack fronts, indicate that the AF-55 does allow for relative uncoupling between the layers. In this respect the adhesively-bonded laminates differ from those used in what is evidently the only other investigation of fatigue crack propagation in through-cracked laminates of divider orientation, that of Plumbridge and Ryder (Ref. 8). Their work, also using aluminum alloys, likewise showed no difference in crack growth rates between monolithic and laminated material. They attributed the lack of improvement in their laminates to a bond strength, produced by hot rolling, which was too high to allow the uncoupling required for crack-divider behavior.

SUMMARY AND CONCLUSIONS

Comparison of fatigue crack growth rates in 1/2 in. thick 7075-T651 aluminum alloy plate material and in laminates of two or four layers made by adhesive bonding shows no significant differences. While in a few cases the laminates exhibited somewhat lower growth rates at low stress intensity values, this trend was not consistent and is attributed to variability in the initial crack sizes formed in the several layers prior to growth rate testing. Fracture surfaces of the laminated panels indicate that the through-the-thickness cracks were able to propagate with some degree of independence, as is desirable in a crack-divider laminate. However, the smallest laminae thicknesses used -- 1/8 in. -- are not sufficiently thin to markedly influence growth rates.

The results demonstrates that adhesively bonded structures can be analyzed with respect to fatigue using the base metal properties alone if through-the-thickness cracks are to be assumed present. It is recommended for design purposes that growth rate data corresponding to the lamina thickness rather than the overall thickness be used when available.

REFERENCES

1. J. A. Alic and H. Archang, "Comparison of Fracture and Fatigue Properties of Clad 7075-T6 Aluminum in Monolithic and Laminated Forms," SAE Paper 750511, 1975.

2. C. E. Feddersen, "Discussion," Plane Strain Crack Toughness Testing of High Strength Metallic Materials, ASTM STP 410, pp. 77-79, 1969.

3. O. L. Bowie, "Analysis of an Infinite Plate Containing Radial Cracks Originating at the Boundary of an Internal Circular Hole," J. Math. Phys. 35, pp. 60-71, 1956.

4. G. T. Hahn and R. Simon, "A Review of Fatigue Crack Growth in High Strength Aluminum Alloys and the Relevant Metallurgical Factors," Engng. Frac. Mech. 5, pp. 523-540, 1973.

5. D. P. Wilhem, "Investigation of Cyclic Crack Growth Transitional Behavior," Fatigue Crack Propagation, ASTM STP 415, pp. 363-383, 1967.

6. D. Broek and J. Schijve, "The Influence of Sheet Thickness on Crack Propagation," Aircraft Engng. 38, pp. 31-33, November 1966.

7. W. J. Plumbridge, "The Influence of Specimen Thickness on Fatigue Crack Propagation in Aluminium and Aluminium Alloys," Quantatitive Relation Between Properties and Microstructure, D. G. Brandon and A. Rosen, eds., Israel Universities Press, pp. 399-405, 1969.

8. W. J. Plumbridge and D. A. Ryder, "A Preliminary Study of the Tensile and Fatigue Properties of Autogenous Aluminium and Aluminium Copper Alloy Laminates," J. Inst. Metals 99, pp. 233-237, 1971.

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

RPV GROUND IMPACT ATTENUATOR SIMULATION

Prepared by:	Charles E. Nuckolls, PhD.
Academic Rank:	Assistant Professor
Department and University:	Department of Mechanical Engineering and Aerospace Sciences Florida Technological University
Assignment:	
(Laboratory)	Flight Dynamics
(Division)	Vehicle Equipment
(Branch)	Recovery and Crew Station
USAF Research Colleague:	R. Harley Walker
Date:	August 29, 1975
Contract No.:	F44620-75-C-0031

RPV GROUND IMPACT ATTENUATOR SIMULATION

by

Charles E. Nuckolls

ABSTRACT

There exist both immediate and long range needs for impact attenuators to facilitate ground recovery of the RPV. A mathematical model of a hypothetical vehicle was conceived to assist in determining the requirements for and evaluating the performance effectiveness of such attenuators. Validation of this mathematical model is discussed. Use of the model in conjunction with shock spectra and with a simulation program called KRASH is discussed.

The problem is complicated by the presence of a horizontal component of velocity whose sense with respect to the RPV cannot be controlled. Several design concepts are suggested for possible solution of the immediate problem. Various configurations of an instantaneously deployable rigid foam are suggested for long term investigation.

INTRODUCTION

The present mode of recovery of Remotely Piloted Vehicles (RPV) is referred to as MARS or Mid-Air Recovery System, in which the parachute retarded vehicle is caught in mid air by a helicopter and carried to a soft landing. This sequence of events, depicted in Figure 1, has been necessary because the RPV sustains severe damage upon impact in parachute retarded ground recovery. To permit ground recovery, an impact attenuator will be required. Therefore, the objective of this study has been the development of an analytical model, based on a typical structural configuration for an RPV, to be used in determining the requirements for and the performance effectiveness of an impact attenuating system.

This differs from most aircraft crashworthiness studies in a fundamental sense. The intention is to prevent all structural damage where as in customary crashworthiness studies the objective is to protect the integrity of the passenger compartment and the structure itself is expendable and by design fails in a controlled manner. Nevertheless, advantage can be taken of the work that has been done in that field. In particular, there exists a simulation program called KRASH [1] which can be adapted to the RPV. This program was developed for helicopter crashworthiness studies by Lockheed California for the U. S. Army Air Mobility Research and Development Laboratory in Fort Eustis, Virginia.

DESCRIPTION OF PROGRAM KRASH

The structure is modeled as a collection of interconnected lumped masses (up to 80). Each mass is allowed six degrees of freedom. The masses are connected internally by beam elements which may be non-linear. Each mass is allowed up to three external springs which radiate outward from the mass point to contact the ground, providing external crash forces - these, too, may be non-linear. Furthermore, it keeps track of kinetic, gravitational potential and elastic potential energies, and the energy dissipated by the internal beam damping mechanism, coulomb friction due to sliding on the surface and plastic deformation of the external springs.

The program represents non-linear internal beam elements in terms of KR (stiffness reduction) factors - each non-linear element requiring a table be input containing up to fifteen pairs of displacement and KR factor. This is illustrated below for an element requiring four such pairs. The KR factor is actually the slope of the load displacement curve relative to the initial slope. The above representation means that the slope remains constant until the displacement reaches x_1 and then decreases uniformly to zero at x_2.

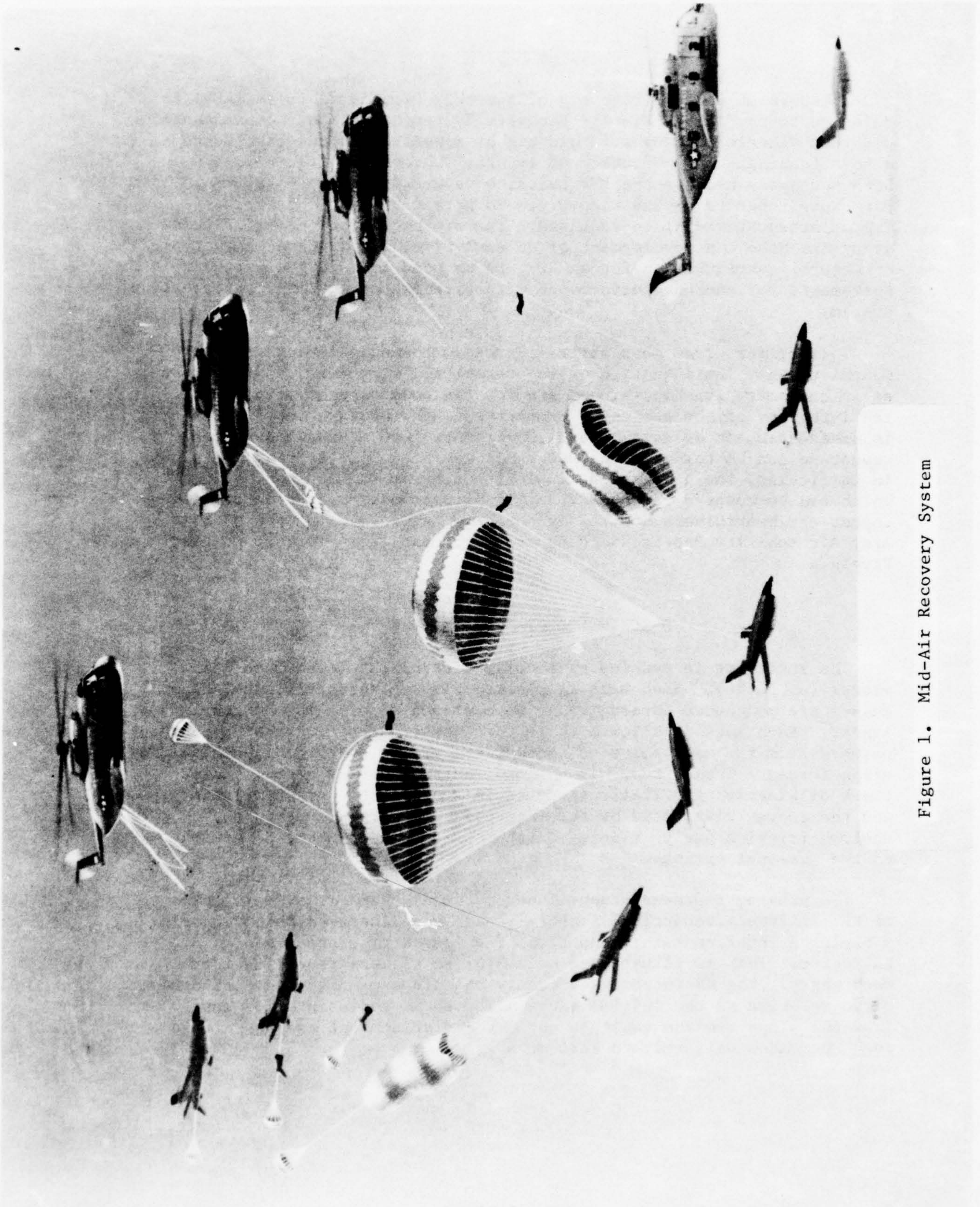
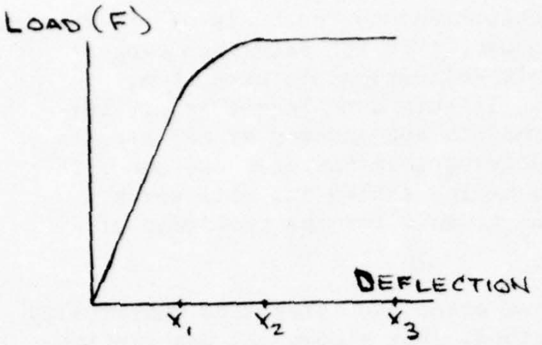
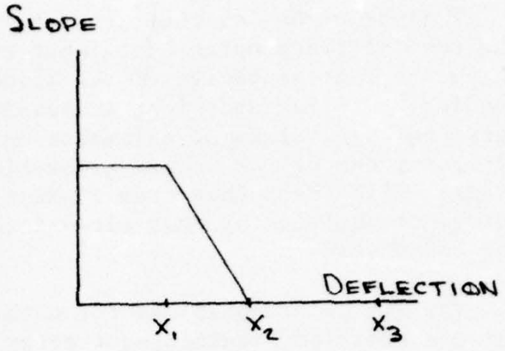


Figure 1. Mid-Air Recovery System





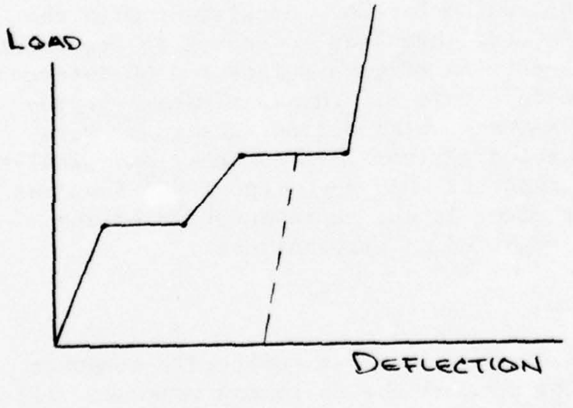
KR TABLE

DEFLECTION	0.0	X_1	X_2	X_3
KR	1.0	1.0	0.0	0.0

The actual load is then calculated according to the scheme:

$$F = K_{initial} \int KR \, dx \qquad (1)$$

The load-displacement curve for the external springs can be represented as generally as shown below:



The coordinates of the break points on this curve plus the bottoming stiffness are input data which are to be adjusted to fit the attenuating device. The unloading stiffness is assumed to be the same as the bottoming stiffness. It is envisioned that this representation will be sufficiently general to model the attenuating device. If not, a separate subroutine will be required.

Failure of an internal element is determined on the basis of deformation. The user of the program must input six quantities for each beam element which are representative of the allowable deflections in extension, bending, twisting and slope in bending. If this data is not input, very large default values of allowable deformation are assumed by the program. After any one of the stated allowable deformations has been exceeded, the program will treat that beam element as having failed and will set all forces transmitted by that element equal to zero for the remainder of the computation.

The six Euler equations for each mass point are integrated numerically using a modified predictor-corrector method. For a complete description of the manner in which the forces and moments are evaluated, the governing equations of motion and the integration scheme, one must study Volume II of USAAMRDL Technical Report 72-72B, May 1973 [2]. KRASH was subsequently modified [3] to facilitate its use by designers, that is, the input format was simplified. The users guide which should be used is the later version in Volume II of USAAMRDL-TR-74-12B, May 1974 [4].

The program KRASH has some other capabilities which were not used in the present study. These are that any mass point may be designated as a DRI element and that the possibility of mass penetration into an occupiable space can be monitored.

The available version of KRASH was written for an IBM system and a major effort was required to convert it for use on the AFFDL computer which is a CDC 6600.

RPV MODEL

A typical RPV structural configuration is shown in Figure 2 which is a cutaway drawing of the XQM-103, without pods. Consistent with the requirements of program KRASH, a model has been conceived to represent a typical vehicle. This model contains 48 mass points and 60 interconnecting beam elements as is represented in Figure 3. This includes a representation of the pods. The parameters which define this model were selected on the basis of information gleaned from various loads analysis reports and weight and balance reports. One such report [5] involves the BQM-34A. Consequently, the model is not representative of any particular vehicle but must be regarded as hypothetical.

Mass Points

The choice of these mass point locations was admittedly somewhat arbitrary but was made for one or more of the following reasons: (1) they are critical points as indicated by failure modes in previous drop tests - such as 4, 5, 13, 14 which are the points of attachment of the nacelle to the fuselage; (2) they are so called "hard points" or points which are easily defined - such as 6, 7, 8 which are engine mounts or 11, 12, 15, 17 which are the wing attachment points and 34, 44, 38, 46, 33, 43, 37, 45 which represent the pylons; (3) loads may be input at these points -

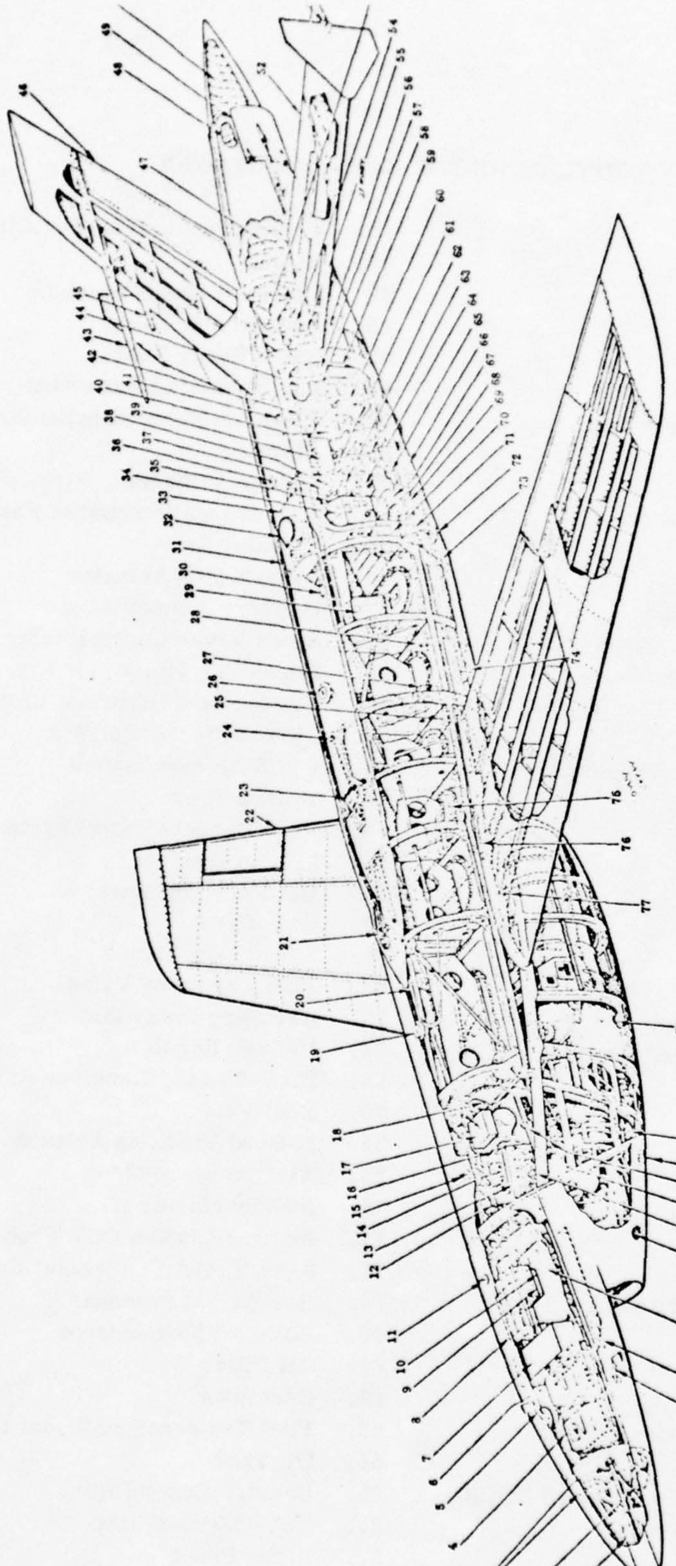


Figure 2. Cutaway Drawing of the XQM-103 Vehicle

KEY LIST/LEGEND FOR XQM-103 CUTAWAY

1. TV Camera
2. Camera Tilt Actuator
3. Ballast
4. Glide Battery
5. Beta Probe
6. Instrumentation Battery
7. Forward Data Distribution Box
8. Compass
9. Heat Sink
10. Instrumentation Transmitter
11. Top Video Antenna
12. Rpm Converter
13. Main Power Control Box
14. Instrumentation Antenna Coupler
15. Top Instrumentation Antenna
16. Instrumentation Battery Disconnect
17. 750 VA Inverter
18. Fuel Flow/Temperature Sensors
19. Fuel Topoff Disconnect
20. Fuel Tank Vent Line
21. Forward Shackle
22. Aileron
23. Aft Shackle
24. Captive Load Fitting
25. Accelerometers
26. Fuel Topoff Vent Disconnect
27. Instrumentation "L" Band Beacon
28. Fuel Dump Valve
29. Fuel Pump
30. Special Devices
31. Parachute Riser
32. Umbilical Connector
33. Distribution Box
34. Flight Control Computer
35. SW-7 Rate Torque Test Switch
36. Top VTCS Antenna
37. Sequential Timer
38. Relay Control Box
39. Instrumentation Signal Conditioner
40. DRW-29/BCR-50A
41. Instrumentation Altitude/Airspeed XDCR
42. Instrumentation PCM Encoder
43. Pitot Tube
44. Mach Transducer
45. Instrumentation Outside Air Temp. Probe
46. Rudder
47. 100-Foot Main Parachute
48. Engagement Chute
49. Speed Brake Door
50. 8.8-Foot Drag Parachute
51. Elevator-Position Indicator
52. Elevator
53. Rudder Actuator
54. Instrumentation Spares Panel
55. Elevator Servo
56. Speed Brake Actuator
57. Delivery Jumper Plug
58. Speed Brake Control Relay
59. Sequential Timer
60. Telemetry Calibration Unit (2)
61. Telemetry Control Box
62. VTCS Antenna Switch
63. Transponder
64. Vega Target Control System Tray
65. 250 VA Inverter
66. Bottom VTCS Antenna
67. Rate Gyro
68. Signal Conditioner
69. Fuel Expulsion Valve
70. Recovery Relay Box
71. Voltage Regulator
72. DRW-29 UHF Receiver Antenna
73. Fuel Vent
74. L-Band Tracking Antenna
75. Aileron Servo (LSI)
76. Accelerometer (2)
77. Instrumentation EGT Probes
78. J-69-T-41A Continental Engine
79. Bottom TM Antenna
80. Bottom Video Antenna
81. Oil Filler
82. Generator
83. Fuel Temperature Signal Conditioner
84. Oil Tank
85. Camera Control Box
86. Video Transmitter
87. Alpha Probe

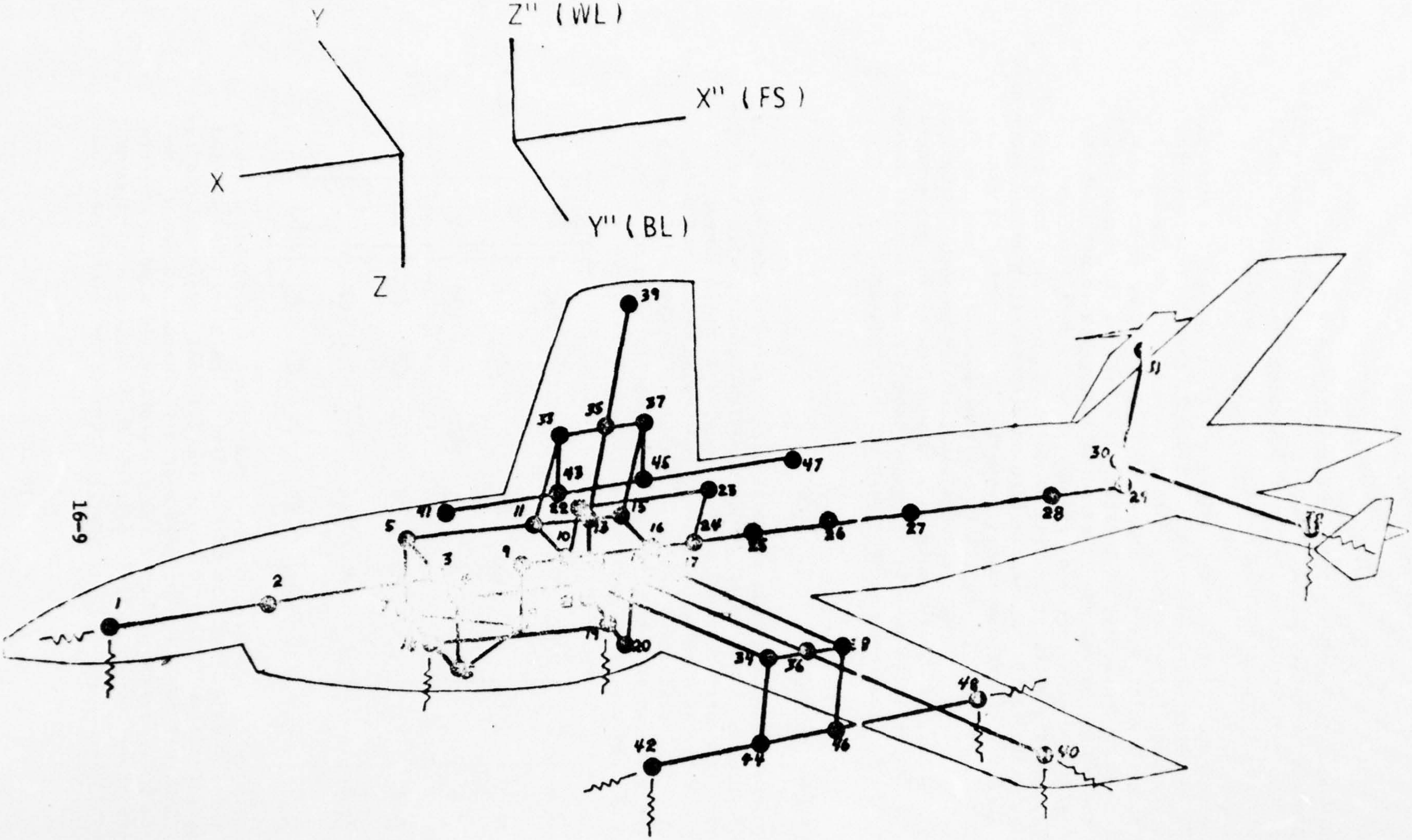


Figure 3. Lumped Parameter Model of RPV

for example 22 and 23 which are forward and aft shackles or 19, 20, 21 which represent the main nacelle crash frame; or (4) their choice would facilitate the calculation of the interconnecting element stiffness such as 18, 19 which represents the skid or keel of the nacelle.

The coordinates of the chosen points are given in Table I according to the (") coordinate system shown in Figure 2, that is Fuselage Station, Butt Line and Water Line. The weights in pounds force of the masses assigned at each point are also listed in Table I. Some adjustments were made to keep the moment of inertia (if known) of the components (such as the pods) correct and to effect a reasonable location of the center of gravity. The center of gravity is located at Fuselage Station 78. Mass moments of inertia in units of in-lb-sec^2 have been assigned as in Table II. These are with respect to right handed coordinate systems fixed in the various masses, parallel to the center of gravity coordinate system fixed in the RPV (x positive forward, y positive right, z positive downward). Those not shown are taken to be zero. Many mass points have been arbitrarily assigned a value of unity for all moments of inertia. These could have in principle been treated as point masses but the KRASH program requires that none of the three moments of inertia be zero.

Interconnecting Elements

Sixty interconnecting beam elements are required by this model, as shown in Figure 2. Since we are concerned with recovery with no structural damage, the beam elements will be modeled as being linear and perfectly elastic with an exception for the engine mounts which can bottom out. The flexibility matrix will thus be determined for each element and inverted to obtain the stiffness matrix. Use of the principle of superposition leads to Equation 2.

$$\begin{Bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta \phi \\ \Delta \theta \\ \Delta \psi \end{Bmatrix} = \begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 & 0 & a_{26} \\ 0 & 0 & a_{33} & 0 & a_{35} & 0 \\ 0 & 0 & 0 & a_{44} & 0 & 0 \\ 0 & 0 & a_{53} & 0 & a_{55} & 0 \\ 0 & a_{62} & 0 & 0 & 0 & a_{66} \end{bmatrix} \begin{Bmatrix} X \\ Y \\ Z \\ L \\ M \\ N \end{Bmatrix} \qquad (2)$$

These quantities are with respect to a beam coordinate system which is fixed in the ith mass, with its origin at m_i. The x axis is directed along a straight line from m_i to m_j in the original undeformed configuration. The roll angle of the beam axes with respect to the m_i axes is input to the program so that y and z are principle axes of inertia of the beam cross section. X, Y, Z and L, M, N are force and moment components acting at point j on beam ij in beam axes. Likewise, the

TABLE I

	MASS POINT	COORDINATES (FS, BL., WL)	WEIGHTS
1.	Nose	(-48.75, 0.0 , 20.0)	122.6
2.	Nose	(- 6.65, 0.0 , 20.0)	150.0
3.	Nose Station 35	(35.0 , 0.0 , 20.0)	50.0
4.	Left Front Nacelle Hanger	(35.0 , 11.77, 20.0)	50.0
5.	Right Front Nacelle Hanger	(35.0 , -11.77, 20.0)	50.0
6.	Left Front Engine Mount	(31.3 , 9.44, 9.68)	20.0
7.	Right Front Engine Mount	(31.3 , - 9.44, 9.68)	20.0
8.	Engine	(40.8 , 0.0 , 9.68)	415.2
9.	Rear Engine Mount	(40.8 , 0.0 , 20.0)	11.6
10.	Wing Carry Through Structure	(66.75, 0.0 , 20.0)	11.6
11.	Right Front Wing Attach & Rear Nacelle Hanger	(66.75, -12.25, 20.0)	23.2
12.	Left Front Wing Attach & Rear Nacelle Hanger	(66.75, 12.25, 20.0)	23.2
13.	Right Crash Frame "V" Block	(76.0 , -12.25, 20.0)	24.5
14.	Left Crash Frame "V" Block	(76.0 , 12.25, 20.0)	24.5
15.	Right Rear Wing Attach Point	(85.0 , -12.25, 20.0)	30.8
16.	Wing Carry Through Structure	(85.0 , 0.0 , 20.0)	30.8
17.	Left Rear Wing Attach Point	(85.0 , 12.25, 20.0)	30.8
18.	Front of Skid (Keel)	(31.3 , 0.0 , 9.68)	20.0
19.	Rear of Skid	(76.0 , 0.0 , 4.19)	20.0
20.	Left Crash Frame	(76.0 , 12.25, 4.19)	20.0
21.	Right Crash Frame	(76.0 , -12.25, 4.19)	20.0
22.	Forward Shackle	(66.75, 0.0 , 38.6)	23.2
23.	Aft Shackle	(99.0 , 0.0 , 37.3)	23.2
24.	Fuselage Station 99	(99.0 , 0.0 , 20.0)	69.2
25.	Fuselage Station 112	(112.0 , 0.0 , 20.0)	92.4
26.	Fuselage Station 130	(130.5 , 0.0 , 20.0)	92.4
27.	Fuselage Station 147	(147.0 , 0.0 , 20.0)	92.4
28.	Fuselage Station 162	(162.7 , 0.0 , 20.0)	92.5
29.	Empennage	(201.0 , 0.0 , 20.0)	29.6
30.	Empennage	(201.0 , 0.0 , 24.0)	29.6
31.	Empennage	(231.0 , -30.0 , 24.0)	30.0
32.	Empennage	(231.0 , 30.0 , 24.0)	30.0
33.	Right Front Wing Spar	(85.5 , -31.0 , 20.0)	24.0
34.	Left Front Wing Spar	(85.5 , 31.0 , 20.0)	24.0
35.	Right Center Wing Spar	(94.75, -31.0 , 20.0)	23.9
36.	Left Center Wing Spar	(94.75, 31.0 , 20.0)	23.9
37.	Right Rear Wing Spar	(103.75, -31.0 , 20.0)	24.0
38.	Left Rear Wing Spar	(103.75, 31.0 , 20.0)	24.0
39.	Right Wing Tip	(136.0 , -72.25, 20.0)	30.0
40.	Left Wing Tip	(136.0 , 72.25, 20.0)	30.0
41.	Pods	(29.0 , -31.0 , 3.6)	20.0
42.	Pods	(29.0 , 31.0 , 3.6)	20.0
43.	Pods	(85.5 , -31.0 , 3.6)	90.4
44.	Pods	(85.5 , 31.0 , 3.6)	90.4
45.	Pods	(103.75, -31.0 , 3.6)	90.4
46.	Pods	(160.25, 31.0 , 3.6)	90.4
47.	Pods	(160.25, -31.0 , 3.6)	20.0
48.	Pods	(160.25, 31.0 , 3.6)	20.0

TABLE II

MASS MOMENTS OF INERTIA

	I_{xx}	I_{yy}	I_{zz}	I_{xy}	I_{yz}	I_{xz}
1.	24.0	12.0	12.0			
2.	24.0	12.0	12.0			
3.	1.0	1.0	1.0			
4.	1.0	1.0	1.0			
5.	1.0	1.0	1.0			
6.	1.0	1.0	1.0			
7.	1.0	1.0	1.0			
8.	65.0	238.0	217.0			-32.0
9.	1.0	1.0	1.0			
10.	1.0	1.0	1.0			
11.	1.0	1.0	1.0			
12.	1.0	1.0	1.0			
13.	1.0	1.0	1.0			
14.	1.0	1.0	1.0			
15.	1.0	1.0	1.0			
16.	1.0	1.0	1.0			
17.	1.0	1.0	1.0			
18.	1.0	1.0	1.0			
19.	1.0	1.0	1.0			
20.	1.0	1.0	1.0			
21.	1.0	1.0	1.0			
22.	1.0	1.0	1.0			
23.	1.0	1.0	1.0			
24.	24.0	12.0	12.0			
25.	24.0	12.0	12.0			
26.	24.0	12.0	12.0			
27.	24.0	12.0	12.0			
28.	24.0	12.0	12.0			
29.	12.0	6.0	6.0			
30.	12.0	6.0	6.0			
31.	3.75	3.75	7.5	-2.0	0.0	0.0
32.	3.75	3.75	7.5	2.0	0.0	0.0
33.	10.0	1.5	13.5			
34.	10.0	1.5	13.5			
35.	10.0	1.5	13.5			
36.	10.0	1.5	13.5			
37.	10.0	1.5	13.5			
38.	10.0	1.5	13.5			
39.	8.4	8.4	16.9	-4.7	0.0	0.0
40.	8.4	8.4	16.9	4.7	0.0	0.0
41.	5.0	10.0	10.0			
42.	5.0	10.0	10.0			
43.	17.0	10.0	10.0			
44.	17.0	10.0	10.0			
45.	17.0	10.0	10.0			
46.	17.0	10.0	10.0			
47.	5.0	10.0	10.0			
48.	5.0	10.0	10.0			

displacements are expressed for point j relative to i in beam axes. The flexibility coefficients a_{pq} will generally be of the following form:

$$a_{11} = \frac{\ell}{AE}$$

$$a_{22} = \frac{\ell^3}{3EI}$$

$$a_{33} = \frac{\ell^3}{3EI}$$

$$a_{44} = \frac{\ell \oint \frac{ds}{t}}{4A^2 G} \quad \text{OR} \quad \frac{3\ell}{G b t^3} \quad \text{OR} \quad \frac{\ell}{GJ}$$

(THIN WALLED TUBE) (OPEN RECTANGULAR) (CIRCULAR SECTION) (3)

$$a_{55} = \frac{\ell}{EI}$$

$$a_{66} = \frac{\ell}{EI}$$

$$a_{26} = a_{62} = \frac{\ell^2}{2EI}$$

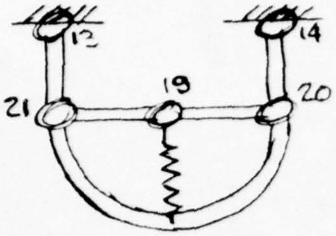
$$a_{35} = a_{53} = \frac{\ell^2}{2EI}$$

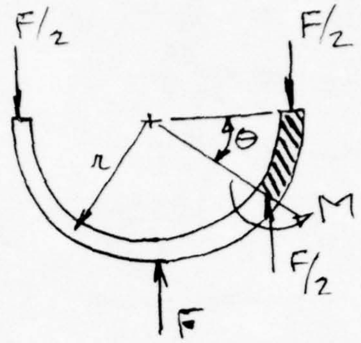
With the exception of a_{11} and a_{44}, these are readily derived by use of the area-moment theorems applied to cantilever beams, where I is the appropriate effective area moment of inertia. The equation used for the torsional flexibility coefficient, a_{44}, depends upon the form of the cross-section. It has been assumed that there is no coupling between bending and twisting.

Sensitivity studies likely would show that the stiffness of some of the elements is not critical and gross errors could be tolerated. It has been assumed that this applies to the following members and their stiffness matrices have been set as an arbitrary large number, 10^{20}, times the identify matrix: 34-36, 36-38, 33-35, 35-37, 10-22, 23-24, 10-11, 10-12, 15-16, 16-17, 11-13, 13-15, 12-14, 14-17.

The response is more sensitive to the stiffness of other elements and a more detailed substructure analysis is required. This is difficult to accomplish analytically and is virtually impossible with the meager information available to date in the form of drawings and reports without further gross assumptions. This was attempted for some of the more critical elements such as the main crash frame and the engine mounts. Their analysis follows and later, in the validation section, suggestions will be made with regard to a testing program for refining the stiffness description.

Main Crash Frame - The main crash frame is represented below along with its lumped parameter model. The external spring at point 19 should represent either the ground or the attenuator system or a series combination of the two. The stiffness of members 19-20 and 19-21 will be adjusted to match that of the actual curved member. At points 13 and 14 are located the "Vee" blocks which are better approximated by pin connections than built in as shown.





Consequently, in the following analysis the curved member will be assumed to be supported as shown in this free body diagram segment of the beam. By the use of Castigliano's principle, $\frac{\partial U}{\partial F} = \delta$, the deflection at the point of application of the force F. From an investigation of equilibrium of the free body, the internal bending moment at position θ is:

$$M = \frac{Fr}{2}(1 - \cos\theta)$$

The strain energy due to bending is now:

$$U = \frac{1}{2EI}\int M^2\,dx$$

By Castigliano's theorem:

$$\delta = \frac{\partial U}{\partial F} = \frac{2}{EI}\int_0^{\pi/2} \frac{Fr^3}{4}(1 - \cos\theta)^2\,d\theta = 0.176\,\frac{Fr^3}{EI}$$

Now beam 20-19-21 (simply supported) is assumed to have stiffness (EI)* such that its displacement under a central force F would be the same as that of the curved beam just computed,

$$\delta = \frac{F\ell^3}{48(EI)^*} = 0.176\,\frac{Fr^3}{EI} \qquad (4)$$

From reference [6], page 4.81, EI for the main crash frame is $3(10^7)\#\text{in}^2$. Thus, the equivalent stiffness for elements 19-20 and 19-21 becomes:

$$(EI)^* = \frac{(EI)(8)}{(48)(.176)} \sim 3(10^7)$$

There is another frame at Fuselage Station 63.5 which is very close to the main crash frame so the two are treated as being in parallel. The frame at station 63.5 is only ~1/10 as stiff as the main crash frame and the stiffnesses assigned to elements 19-20 and 19-21 is therefore consistent with an EI of $3.4(10)^7$.

Engine Mounts - Very little information is available at the present time about the engine mounts except for their form and approximate dimensions which are given in Teledyne Ryan Aircraft drawings 124P302 and 124P304, partially reproduced here as Figure 4. In accordance with information from *Rubber Springs Design* by Gobel [7], the following spring rates were calculated.

The front mounts are sleeved rubber springs in axial shear, shaped so that the nominal shearing stress is uniform. The axial and the radial spring rates are given by equation 2.3 (page 34) and equation 2.20 (page 60) respectively:

$$K_{axial} = \frac{2\pi r l G\,[1]}{r_2 - r_1} = 3.38(10^3)\ \#/in$$

$$K_{radial} = \frac{7.5\pi h G}{\ln(r_2/r_1)} K_1 = 1.23(10^4)\ \#/in$$

Where k_1 is a form factor determined from Figure 2.22, page 60.

These stiffnesses apply for deflections roughly as large as 34% of the corresponding dimension of the mount, page 33. These are 0.54" axially and 0.15" radially after which the stiffness will be assumed to be, arbitrarily, three times as great.

The aft engine mount is essentially a steel tensile link, refer to drawing 124P304. The stiffness in the axial direction is:

$$k = \frac{AE}{L} = 2.36\ (10^6)\ \#/in$$

All other stiffnesses for the aft engine mount are zero by virtue of the spherical joint.

Allowable Deformation - Numbers for the allowable elastic deflections have been determined, resulting from very simple beam analyses, and some more gross assumptions. In this situation, exceeding the elastic limit in a member does constitute failure but not rupture and the forces should not be returned immediately to zero. Therefore the members are modeled as being ductile and having the load displacement properties discussed earlier. The numbers that describe the plastic portion of the element load displacement curves have been arbitrarily set to an integer multiple of the allowable elastic deformation. One disadvantage of all this is that the computer will print out an indication of failure after complete rupture whereas for our purposes failure occurs when the elastic limit is exceeded.

[1] Assuming that the rubber hardness is 70 IRHD, the shear modulus, G, is 1.34 Newtons/square millimeter, Figure 1.11, page 25.

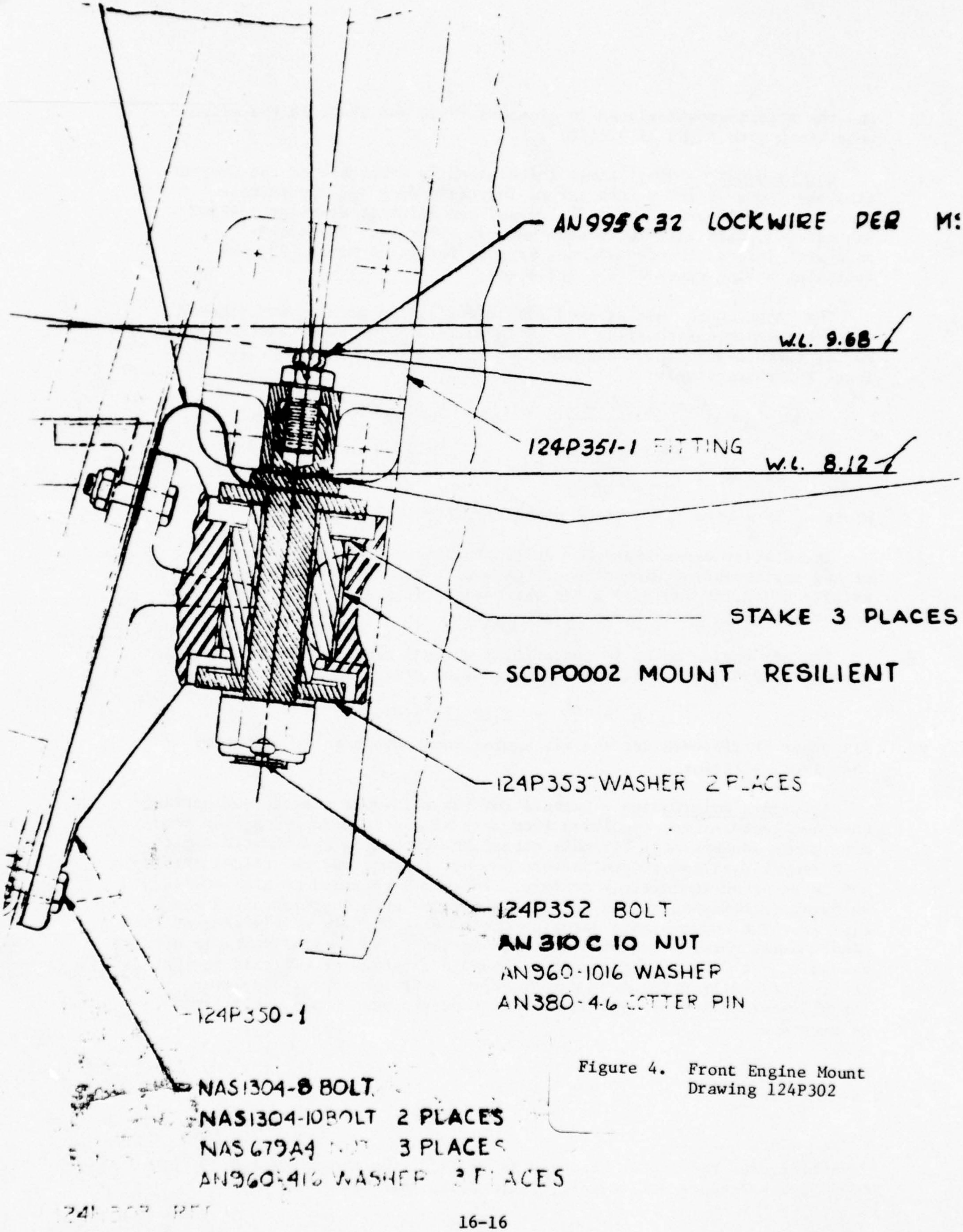


Figure 4. Front Engine Mount
Drawing 124P302

It should also be noted that the actual stress state is one of combined stress and a failure theory such as the maximum energy of distortion should, in principle, be used. Some combination of forces, in other words, should be used instead of a single force as the failure criterion. However, in view of the other assumptions which have been made and the difficulty in implementing this, it is not recommended.

MODEL VALIDATION

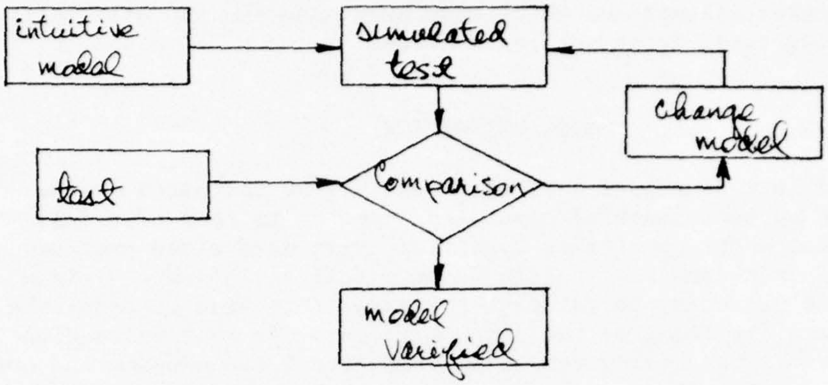
The model just developed has, of course, to be validated. This could be done by comparison of simulated response to that of a full scale drop test under controlled conditions with hard wired instrumentation. Such a drop test should be on a full scale vehicle since the purpose is primarily to validate the parameters used to model the vehicle and not the logic of the program itself. Or this validation could conceivably be done by computing the natural frequencies and mode shapes and comparing these with experimentally determined resonant frequencies and mode shapes. Since time did not permit either of these, an attempt was made to duplicate the failure modes that had been observed in an earlier series of flight tests on the RPV. The primary interest in these tests was on the parachute system, but the impact conditions are known approximately and vehicle damage has been well documented photographically.

The test chosen for the validation was the multi-canopy recovery system third hulk test, conducted 1 May 1975 at Dugway Proving Grounds, Utah. The impact conditions were (i) no angular velocity, (ii) level attitude, (iii) vertical velocity of 17.4 fps, horizontal velocity of 13 fps with a heading of 7:30 with respect to the longitudinal axis (12:00 forward).

The external springs represent the ground in this case. Nothing quantitative was known about the spring rate of the soil so, once again, some gross assumptions were made. Measurements of static load penetration for 6" radius spheres into compacted soil were made at the University of Texas [8] in 1963. It was assumed that this spherical data could be scaled up directly proportional to radius and that it could be extrapolated to much larger penetrations and, further, that a simple fudgé factor could be applied to convert to cylindrical bodies. This has neglected inertial and viscous properties of the soil. It is quite possible that a more extensive literature search would yield information directly for cylindrical bodies. Such information is needed only for this validation attempt and not when a cushion is being modeled. Admittedly the compliance of the ground is in series with that of the cushion but the effective spring rate of the series combination should be for all practical purposes that of the cusnion regardless of the type of terrain at the point of impact.

Model Refinement

In a previous section it was stated that model verification should be done. It was implied that this might be accomplished by a "direct verification" as depicted below. The difficulty with this approach is that the changes to be made in the model are intuitive, no indication of what changes to be made being given.



The success of this technique depends upon the skill of the analyst and the complexity of the model and does not appear to be encouraging. There has been a great deal of interest recently in the direct identification of structures from dynamic test data, see for example the recent ASME monograph [9]. It is not felt that the state-of-the-art of direct system identification has advanced to the point that its use should be attempted to define the abstract (i.e., not measurable) quantities of masses and stiffness required in this model. On the other hand, modal techniques might be applicable if used in conjunction with shock spectra - a technique which is discussed later in this report.

The authors of program KRASH have concluded after extensive tests of helicopter airframe structure that static tests provide satisfactory load-deflection data such as peak failure load, failure deflection, energy absorption and curve shape when compared to similar data obtained from dynamic tests. The RPV, like the helicopter, is a lightweight structure in which failure would be associated with a lower mode and a static analysis, although slightly conservative, should suffice. In addition, a static test requires a lesser amount of instrumentation, allows more rapid test set-up and is consequently more economical.

It is recommended that a static test program be undertaken to determine the stiffnesses, failure loads and failure deflections of those elements which have shown to be critical in prior drop tests - in particular, the wing spars, the forward nacelle attachment points, the rear nacelle attachment points, the crash frame and the pylons. On the other hand, much of this information for the RPV may be in various Teledyne Ryan reports, if available. Ultimately, sensitivity studies may have to be performed to determine what the affect of inaccuracies in the model parameters is upon the overall behavior.

If one would then like to demonstrate the validity of the model a direct dynamic verification of the type mentioned could be attempted.

USE OF SHOCK SPECTRA

In the method just discussed, failure occurs as the result of overdistortion of one or more parts of the structure. And as one might conclude from that discussion, the analytical determination of the maximum distortion in an exact sense is not a simple computation. However, if the shock spectra of the excitation are available and if the motion of the structure is assumed to be linear and to possess classical normal modes, a theoretical upper bound on any element distortion can easily be determined by modal superposition [10]. We have already assumed that the structure is linear - referring to the RPV and not including the attenuating system and that the normal modes exist. The difficulty is that the shock spectra are not available for the shocks which might be applied to the RPV by the attenuating system. It did not initially appear to be likely, but if it should turn out that the form or shape of the shock and its duration are predictable, then the use of shock spectra in a dynamic analysis of the RPV might be advantageous. For that reason, a brief discussion of the theoretical background of shock spectra will be included here. The governing equations of motion of the structure are of the form:

$$\left[m \right]\left\{ \ddot{z} \right\} + \left[k \right]\left\{ z \right\} = -\left\{ \ddot{y} \right\}$$

Where the mass and stiffness matrices are n x n (in the proposed model, n = 288), $\{z\}$ are the relative displacements or distortions, and $\{y\}$ is base motion. Solution of the eigenvalue problem ($\{\ddot{y}\} = 0$) yields the natural frequencies and mode shapes which allows the formation of the coordinate transformation

$$\left\{ z \right\} = \left[q \right]\left\{ \eta \right\} \tag{5}$$

where the square matrix [q] consists of the eigenvectors, normalized so that $\langle q \rangle^{r} [m] \{q\}^{s} = \delta_{rs}$ and $\{\eta\}$ are principle coordinates. Use of this transformation will uncouple the equations of motion, the form of which will be:

$$\left\{ \ddot{\eta} \right\} + \left\lceil \omega_n^2 \right\rfloor \left\{ \eta \right\} = -\left[q \right]^{T}\left\{ \ddot{y} \right\}$$

In other words, each equation of motion is now of the form:

$$\ddot{\eta}_i + \omega_i^2 \eta_i = -Y(t) \tag{6}$$

Where the function $Y(t)$ represents the shock excitation to the system.

The solution of (6) is given by the Duhamel integral

$$\eta_i(t) = -\frac{1}{\omega_i}\int_0^t Y(\tau)\sin\omega_i(t-\tau)\,d\tau \tag{7}$$

Equation (5) then in summation form is

$$z_r(t) = \sum_{i=1}^{n} \alpha_{ri}\eta_i(t)$$

Where the α_{ri} may be regarded as mode participation factors.

The maximum absolute value of $z_r(t)$, denoted as ZMAX, is defined as:

$$\text{ZMAX} = \max_{t>0}\left|z(t)\right| = \max_{t>0}\left|\sum_{i=1}^{n}\alpha_{ri}\,\eta_i(t)\right|$$

An upper bound on ZMAX, denoted as ZU, is:

$$\text{ZU} = \sum_{i=1}^{n}\left|\alpha_{ri}\right|\left|\eta_i\right|_{max}$$

The computation of the $|\eta_i|_{max}$ for the structure would be as difficult as finding ZMAX, except that for a given shock the $|\eta_i|_{max}$ can be determined by reference to the shock spectrum. The determination of the shock spectrum, of course, requires repeated solution of equation (7) but this can be carried out separately.

It is noted that this approach requires the solution of the eigenvalue problem, but only once. However, for the model which has been developed, the number of degrees of freedom is 48 x 6 = 288 and such solution would be next to impossible. The model would have to be simplified while retaining enough detail to describe distortion in those elements which are critical - such as the wing spars, the crash frame, the nacelle hangers, the pylons, etc. Once this has been done, it would be relatively easy to compute an upper bound on the distortion of any element.

This procedure is not without merit but is somewhat abstract and was discarded in favor of the more straight forward appraoch offered by KRASH.

DESIGN CONCEPTS

There exist both long range and immediate applications for impact attenuating devices. Existing air bag technology could be applied for a solution of the immediate RPV problem if it were not for the presence of a horizontal component of velocity. This horizontal component approaches in magnitude that of the vertical component. Furthermore, the orientation of this component with respect to the longitudinal axis cannot be controlled. The energy associated with this horizontal velocity can best be dissipated through friction.

Air bags cannot sustain a shearing type of force which would result from sliding on the surface whereas a rigid foam can. Rigid foams also have more flat load displacement profiles than do air bags which means that a smaller stroke or deformation is required. A rigid foam is thus better suited for this application than an air bag but has obvious problems associated with deployment.

There are foams available today for which it is said deployment is virtually instantaneous, and full mechanical properties are developed within seconds. No literature on such foam is available but their existence opens up great possibilities.

Some suggestions are made with regard to an immediate fix but all are deficient with respect to the vehicle overturning, particularly if the horizontal component of velocity is transverse.

Immediate

There is apparently an immediate need for a shock attenuator for the AQM-34V vintage model RPV presently in use. A study has been conducted [11] which indicates that the damage sustained in ground recovery by the AQM-34H is much more severe than that of the BQM-34A. The significant difference between the two is weight and the presence of the pods on the AQM-34H. This suggests an obvious action - jettison the pods. Or jettison the pods but restrain them on a tether. This would be further beneficial in that the tethered pods would drag along the ground, dissipating a bit of energy and orienting the RPV into a more favorable longitudinal attitude prior to impact. This procedure could be easily implemented and should reduce the damage sustained on impact to the levels experienced with the BQM-34A. Whether or not this would be compatible with existing operating procedures cannot, of course, be determined by this author.

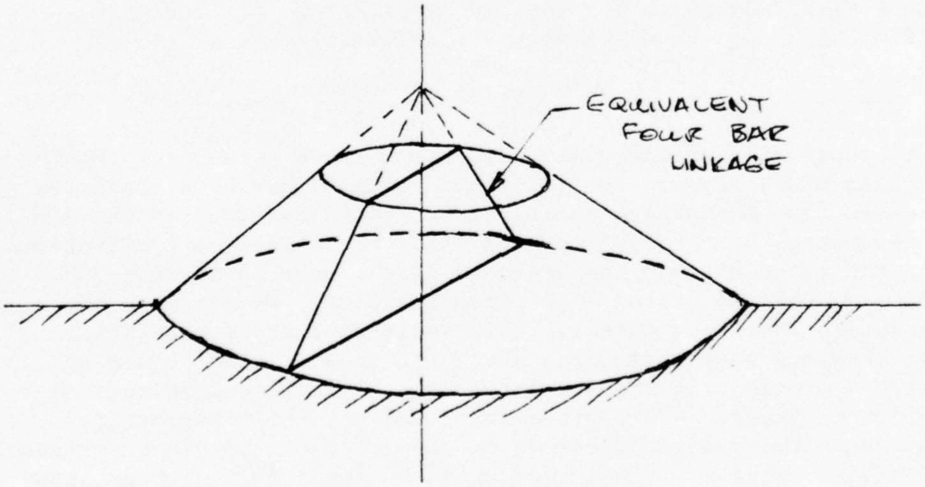
The pods protrude below the nacelle and would contact a smooth flat surface first. The pod and wing combination must then deflect significantly before the nacelle contacts the ground. It has been observed that the pods sustain little permanent damage whereas the wing root is heavily damaged. This suggests that redesign of the pylons as frangible elements would be beneficial, but of limited value because the stroke of the crushable element would be limited. No kinematic schemes for accomplishing this are suggested, but a large body of technology exists since the recent interest in energy absorbing automobile bumpers.

One concept using an instant foam which might be feasible for the present generation RPV is to deploy two cylindrical bodies of foam - one on each side lodged between the nacelle and the pod. These would be sized so that upon impact they would be extruded up into the space between the nacelle and pod, which is roughly seven inches in width. This concept is suggested as one of the easier to implement by use of an instant foam or by filling a bag tightly with foam spaghetti, recognizing its deficiency with respect to the vehicle overturning.

Other design schemes were ruled out because they would have involved extensive retrofitting of hardware or extensive development.

Long Range

Recognizing the omni-directional aspects of the problem, a conical configuration comes to mind. If the problems of deployment of a shallow conical shell foam device can be solved, this concept could be an elegant solution. If one large truncated cone were used, the action would likely be similar to that in a four bar linkage. The four bar linkage is a quite non-linear device but whether or not this non-linearity will work to advantage and could be exploited will require further investigation.



The cone half angle (β) should be large enough so that slipping will occur before overturning, unless an obstacle is encountered, in which case local crushing of the rim of the cone will absorb some if not all of the energy. In order to preclude overturning, the cone half angle should satisfy the relation

$$\beta \geq \tan^{-1} \mu$$

where μ is the friction coefficient. Even if the friction coefficient were as big as unity the half angle required would be 45° which is a reasonable value. This concept is quite avant-garde but seems to be worthy of further study.

RECOMMENDATIONS

The parameters which have been set down represent a hypothetical vehicle. These were based on some rather extensive idealizations of the structure. It would be best if these parameters could be tailored to represent more precisely the RPV, but even if this cannot be achieved the model will be useful for comparison of cushioning schemes in a qualitative sense.

Almost certainly a static testing program will be required to effect this tailoring of the model. The stiffnesses and allowable forces for the critical elements must be measured. The most critical elements seem to be the wings, pylons, and nacelle attachment fittings.

The overall vehicle response will be more sensitive to errors in some parameters than in others. These sensitive parameters should be located by repeated use of the program, and their load-deflection characteristics measured statically as well.

After a confidence in the model has been established, parameter studies should be undertaken. In particular, the affects of nose-up vs. nose-down attitudes must be assessed and allowable limits for horizontal velocity determined.

REFERENCES

1. Wittlin, G., M. A. Gamon, *Computerized Unsymmetrical Mathematical Simulation and Experimental Verification for Helicopter Crashworthiness in Which Multidirectional Impact Forces Are Present*, USAAMRDL Technical Report 72-72A - Experimental Program for the Development of Improved Helicopter Structural Crashworthiness Analytical and Design Techniques, Volume I, May 1973.

2. Wittlin, G., M. A. Gamon, *Test Data and Description of an Unsymmetrical Crash Analysis Computer Program, Including a User's Guide and Sample Case*, USAAMRDL Technical Report 72-72B, Volume II, May 1973.

3. Wittlin, G., K. C. Park, *Development of Simplified Analytical Techniques to Predict Typical Helicopter Airframe Crushing Characteristics and the Formulation of Design Procedures*, USAAMRDL-TR-74-12A - Development and Experimental Verification of Procedures to Determine Nonlinear Load-Deflection Characteristics of Helicopter Substructures Subjected to Crash Forces, Volume I, May 1974.

4. Wittlin, G., K. C. Park, *Test Data and Description of Refined Program "Krash", Including a User's Guide and Sample Case*, USAAMRDL-TR-74-12B, Volume II, May 1974.

5. *Weight and Balance Report, BQM-34A*, Report No. 12444-49, Teledyne Ryan Aeronautical, San Diego, California, December 1969.

6. *Nacelle Structural Analysis USAF XQ-2C Target Drone*, Report No. 12442-6, Ryan, April 1959.

7. Göbel, E. F., *Rubber Spring Design*, John Wiley and Sons, Inc., New York, 1974.

8. Reese, L. C., et. al., "Investigation of the Effects of Various Soil Conditions on the Landing of a Manned Spacecraft", The University of Texas, Structural Mechanics Research Lab, Austin, Texas, December 1963.

9. *System Identification of Vibrating Structures*, ASME, New York, 1972.

10. Butzel, L. M. and H. C. Merchant, "The Use and Evaluation of Shock Spectra in the Dynamic Analysis of Structures", ASME Paper No. 73-APM-22.

11. McHugh, A. P., "Operations Evaluation of TAC Drone Ground Recovery", USAF Tactical Air Warefare Center, Eglin AFB, July 1973.

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT-PATTERSON AFB, OHIO

&

EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

A METHOD FOR INVESTIGATING THE ANGULAR VIBRATION RESPONSE OF A STRUCTURE

Prepared by:	Philip C. Rymers Ph.D.
Academic Rank:	Professor Mechanical Engineering University of Nevada, Reno
Assignment:	
(Laboratory)	Air Force Flight Dynamics Laboratory
(Division)	Vehicle Dynamics Division
(Branch)	Aerospace Dynamics Branch
USAF Research Colleague:	R. N. Bingman
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

A METHOD FOR INVESTIGATING THE ANGULAR VIBRATION RESPONSE OF A STRUCTURE

by

Philip C. Rymers

ABSTRACT

At any station on an aircraft in flight, the structure is subject to certain input forcing functions. The response to these functions may be such that not only linear but also angular displacements may occur. The input forcing functions are random in nature and result from acoustic pressures produced by engines, aerodynamic loads, buffeting, turbulence and others. Historically only linear motions have been of substantial interest. Due to current interest in optical devices stored on aircraft, possibly at remote locations, the angular motions interest has arisen.

This work consists of attempting to analytically determine if the data banks of linear vibration information can be used to determine these angular motions.

Beyond this effort, the mathematical model could perhaps be refined and testing of the hypothesis performed using available aircraft data.

ACKNOWLEDGEMENTS

The author acknowledges the assistance and encouragement given by J. Fred O'Brien, Jr., Auburn University, the program director, Ralph Bingman, Vibrations Group leader, Karen Hall, and Lt. W. R. Davis of the Vibrations Group who reviewed this report.

LIST OF FIGURES

LIST OF TABLES

PRINCIPLE NOMENCLATURE

EI	beam bending stiffness
w	lateral displacement
m	mass per unit length
W	mode shape
T	time dependent solution
ω	natural circular frequency
ζ	frequency function
ℓ	beam length
y	axial beam coordinate
A,B,C,D,E,F A'_1, B'_1	constants
α_i, β_i	functions of $\zeta_i y$
η_i	function of $\zeta_i \ell$
γ_i	bound on slope for free-free beam
κ	separation constant
J_p, I_p	Bessel functions

SECTION I

Introduction

The mounting of external and internal optical devices on an aircraft airframe so as to minimize degradation of optical resolution leads to an investigation of angular motions of the airframe. In particular, if these optical devices consist of pairs of related components mounted some distance apart, the relative angular motions in a vibration environment may become important.

At a position on the airframe, which is subject to aerodynamic or sonic forcing functions, the response to the input may be such that linear as well as angular motions will result. In general the forcing functions are random in nature and thus the response to these inputs is also random. The nature of the response is such that it may be assumed to be stationary and ergodic without serious error.

This report is addressed to the problem of angular vibration such that the question to be answered may be stated as, "Given the linear vibration response of a structure where amplitude and frequency information about the response is known, can the angular motion response of the structure be deduced?"

The goal of this study is then to find a mathematical formulation of the relationship existing between linear vibration amplitude, frequencies contributing to this amplitude, and the angular vibration amplitude at some point on a structure.

SECTION II

The Slender, Uniform, Free-Free Bar

Initially, in order to maintain the focus of this investigation on the question that has been posed, the aircraft fuselage will be modeled as a slender, uniform, free-free beam. It is intended to eventually extend this model to a more realistic approximation of the aircraft structure.

The differential equation of motion, ignoring rotary inertia and transverse shear components, for a beam is given as [1]

$$(EI\, w'')'' + m\, \ddot{w} = 0 \tag{2.1}$$

where the prime denotes differentiation with respect to beam length y and the dot denotes the time derivative.

Assuming EI and M are constant for all time and positions, the solution to (2.1) may be taken as

$$w = W(y)\,T(t) \tag{2.2}$$

where $$T = A \sin \omega t + B \cos \omega t \tag{2.3}$$

and $$W = C \sinh \xi y + D \cosh \xi y + E \sin \xi y + F \cos \xi y \tag{2.4}$$

In equations (2.3) and (2.4) we have

$$\xi = \sqrt{\frac{\omega}{a}} \qquad a^2 = \frac{EI}{m} \tag{2.5}$$

[1] For derivation of this equation, see Appendix A.

For the free-free beam the boundary conditions are

$$W''(0) = W''(\ell) = W'''(0) = W'''(\ell) = 0 \qquad (2.6)$$

which conditions represent zero bending moment and shear force at the beam ends.

Combining equations (2.4) and (2.6) gives, after some rearrangement of terms

$$\begin{vmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ \sinh \zeta\ell & \cosh \zeta\ell & -\sin \zeta\ell & -\cos \zeta\ell \\ \cosh \zeta\ell & \sinh \zeta\ell & -\cos \zeta\ell & \sin \zeta\ell \end{vmatrix} = 0 \qquad (2.7)$$

which reduced to

$$\cos \zeta\ell = \frac{1}{\cosh \zeta\ell} \qquad (2.8)$$

The solutions to (2.8) give the natural frequencies for the free-free beam as partially tabulated in Table 1.

i	1	2	3	4	'''
$\zeta_i\ell$	1.5056π	2.4998π	3.5π	4.5π	'''

Table 1. Free-Free Beam Natural Frequencies

Continuing with equations (2.4) and (2.6) there results finally

$$W = D\left\{\left[\frac{\cos\zeta\ell - \cosh\zeta\ell}{\sinh\zeta\ell - \sin\zeta\ell}\right](\sinh\zeta y + \sin\zeta y) + (\cosh\zeta y + \cos\zeta y)\right\} \tag{2.9}$$

The slope of the beam at any position y along its length is given by

$$W' = D\left\{\left[\frac{\cos\zeta\ell - \cosh\zeta\ell}{\sinh\zeta\ell - \sin\zeta\ell}\right]\zeta(\cosh\zeta y + \cos\zeta y) + \zeta(\sinh\zeta y - \sin\zeta y)\right\} \tag{2.10}$$

Equations (2.4) and (2.10) give the displacement and slope modes for the free-free beam for one value of $\zeta_i\ell$. For a problem where the solution (2.3) is combined with (2.9) and (2.10) and where many modes are excited we have

$$w = \sum_{i=1}^{\infty}(A_i' \sin\omega_i t + B_i' \cos\omega_i t)\left[\eta_i(\sinh\zeta_i y + \sin\zeta_i y) + (\cosh\zeta_i y + \cos\zeta_i y)\right]$$

$$= \sum_{i=1}^{\infty} f_i(t)\, \alpha_i(\zeta_i y) \tag{2.11}$$

and

$$w' = \sum_{i=1}^{\infty} f_i(t)\, \beta_i(\zeta_i y) \tag{2.12}$$

where in (2.11) and (2.12) we have

$$\eta_i = \frac{\cos \xi_i l - \cosh \xi_i l}{\sinh \xi_i l - \sin \xi_i l} \tag{2.13}$$

and

$$A_i' = AD \quad , \quad B_i' = BD \tag{2.14}$$

Next the ratio of equations (2.12) and (2.11) is formed to give

$$\frac{w'}{w} = \frac{\sum_{i=1}^{\infty} f_i(t)\,\beta_i(\xi_i y)}{\sum_{i=1}^{\infty} f_i(t)\,\alpha_i(\xi_i y)} \tag{2.15}$$

Now if the excitation frequency of the beam occurs very close to one of the natural frequencies, say where i = r, one term, the r^{th} term will dominate in (2.15) and we have

$$\frac{w'}{w} \doteq \frac{\beta_r(\xi_r y)}{\alpha_r(\xi_r y)} \tag{2.16}$$

If the excitation is of sufficiently high frequency so that the modal density of the response is very high then we have

$$\frac{w'}{w} = \frac{\sum_{i=1}^{\infty} \beta_i(\xi_i y)}{\sum_{i=1}^{\infty} \alpha_i(\xi_i y)} \tag{2.17}$$

These results are applicable to sinusoidal response or random response if in the random response problem the mean square displacements and slopes are used in (2.17).

At this point we can conclude that, given the free-free beam displacement and the frequencies contributing to this displacement, the beam slope can be determined by (2.17) since all terms in the right hand side of (2.17) depend on the frequency ζ_i and position y only.

So far in Section II, the slope displacement relation has been formulated for single frequency response and for high modal density or high frequency response.

Consider now the case where the displacement response function has a countable number of peaks as illustrated in Figure 3. The implication of such a response curve is that the natural modes most closely associated with the frequencies f_i, f_j, f_k are making the greatest contribution to the displacement. In fact all other modal contributions might be considered to be minor in comparison.

As a consequence of this, consider only these excited modes in the denominator of Equation (2.15). The sum is then countably finite.

The associated slope function will contain corresponding terms which will dominate in the same manner. If the slope is considered to be the result of applying a linear operator to the displacement, then the above statement follows. Furthermore, the coefficients of the displacement terms are unaltered by the derivative since the independent variable of differentiation is carried by the mode shape terms, not by these coefficients.

By this devise then, the dominant terms of the displacement function and their derivatives are identified and become calculable. Therefore the slope is determined by knowing the effected frequencies, the mode shapes and the displacement for this narrow band problem, correct to within the accuracy of the discarded terms.

SECTION III

Bounds On The Slope of a Free-Free Bar

While the results of Section II show that an affirmative answer to the question prompting this effort may be indicated, a degression is in order at this point.

The maximum possible relative motion between any two points on the beam may be of interest as a consequence of the nature of the general angular vibration problem.

It would appear that the greatest slope (numerically) on the beam would occur at either the ends of the beam or at the node points where the curved (vibrating) beam crosses its equilibrium (rest) position.

To investigate the end point slope we use equation (2.10) with y = 0. This gives

$$\frac{W'(0)\ell}{D} = \zeta_i\ell\left\{ \eta_i\left(\cosh(0) + \cos(0)\right) + \left(\sinh(0) - \sin(0)\right)\right\} \qquad (3.1)$$

$$= 2\zeta_i\ell\eta_i$$

Numerical values of this slope function may be determined using the natural frequencies found from Equation (2.8), partially tabulated in Table 1. Equation (3.1) is greatly simplified however by observing that for values of $\zeta_i\ell>3$, cosh $\zeta_i\ell \rightarrow$sinh $\zeta_i\ell$ so that for practical calculations we have

$$|\eta_i| = 1 \qquad\qquad \zeta_i\ell > 3 \qquad (3.2)$$

Thus for all values of   we can replace (3.1) by

$$\left|\frac{W'(0)\ell}{D}\right| = 2\zeta_i\ell \qquad (3.3)$$

To find the slope at the node points, the inflection points of equation (2.10) must be found. Thus from (2.10) we form

$$\frac{dW'}{dy} = 0 = \eta_1 (\sinh \zeta_1 y - \sin \zeta_1 y) + \cosh \zeta_1 y - \cos \zeta_1 y \tag{3.4}$$

Equation (3.4) is satisfied for all η_i, ζ_i when $y = 0$ and from symmetry (or asymmetry for even modes), when $y = \ell$.

Using equation (3.2), equation (3.4) becomes

$$e^{-\zeta_1 y} = \cos \zeta_1 y - \sin \zeta_1 y \tag{3.5}$$

where use is made of the identity

$$e^{-\zeta_1 y} = \cosh \zeta_1 y - \sinh \zeta_1 y \tag{3.6}$$

The slope at those values of $\zeta_i y$ which satisfy (3.6) will give, along with equation (3.3), the desired bounds of the slope. Note in particular that the bounds for the higher frequencies are given by

$$\gamma_1 = \sqrt{2}\, \zeta_1 \ell \tag{3.7}$$

This result is evident upon examination of equation (2.10) where it must be remembered that $\eta_i \rightarrow -1$ and $\cosh \zeta_i y \rightarrow \sinh \zeta_i y$ for $\zeta_i y > 3$.

SECTION IV

Bars of Varying Stiffness and Mass

A more realistic model of the aircraft stiffness and mass distribution is a second order, skewed polynomial. Of course this is still a model and is not known to be an accurate description of any particular aircraft fuselage.

Appendix B contains a derivation of the distirbution of stiffness as illustrated in Figure 1 at the end of this report.

The results of this exercise are that the stiffness and mass vary as

$$\begin{aligned} EI(y) &= Ay^2 + B \\ m(y) &= Cy^2 + F \end{aligned} \qquad (4.1)$$

No confusion should result here from reusing the constants A, B, C and F.

Returning now to equation (2.1) and using (4.1) there results

$$\frac{d^2}{dy^2}\left[(Ay^2+B)\frac{d^2w}{dy^2}\right] + (Cy^2+F)\frac{d^2w}{dt^2} = 0 \qquad (4.2)$$

which when expanded becomes

$$(Ay^2+B)\,w^{\overline{IV}} + 4Ay\,w''' + 2Aw'' + (Cy^2+F)\,\ddot{w} = 0 \qquad (4.3)$$

Using the separation of variables method as in Section II there ultimately results

$$(Ay^2+B)W^{\overline{IV}} + 4Ay''' + 2AW'' = kW(Cy^2+F) \tag{4.4}$$

and

$$\ddot{T} + kT = 0 \tag{4.5}$$

Comparing (4.5) to (2.3) we see that again the separation constant k plays the familiar role of the frequency.

Equation (4.4) is a linear, fourth order differential equation with variable coefficients which has at this time no known general solution in closed form. Existance of a solution is however known[1].

[1] See ref. 5 page 133.

SECTION V

A Special Case of the Problem of Section IV

J. W. Nicholson has reported in reference 7 on the solution of the problem of lateral vibrations of a free-free beam with stiffness and mass varying as Ay^n where A is a constant and n varies from 0 to 1. The beam is considered to be of circular cross section.

In particular the complete solution for n = 1, the conical rod, is given. This problem differs from that of Section IV in two principle ways. First, the beam is symmetrical about its mid-point and secondly each half of the beam is a solid of revolution of a straight line (Figure 2).

The details of the solution of this problem will not be presented here. However the displacement, a function of Bessel Functions, is given by Nicholson to be

$$W = \frac{1}{z}\left[B\,J_2(2\sqrt{z}) + C\,I_2(2\sqrt{z})\right] \tag{5.1}$$

where

$$Z = y\,(4\,m\omega^2/EA^2)^{\frac{1}{2}}$$

J_2 = Bessel function of the first kind of order 2

I_2 = Bessel function of imaginary argument of order 2

B,C = Constants

The slope of the beam is found from (5.1) to be

$$W' = \frac{1}{z^{3/2}}\left[-B\,J_3(2\sqrt{z}) + C\,I_3(2\sqrt{z})\right] \tag{5.2}$$

and the ratio of (5.2) to (5.1) becomes

$$\frac{W'}{W} = \frac{1}{Z^{1/2}} \frac{\left[-B\,J_3(2\sqrt{Z}) + C\,I_3(2\sqrt{Z})\right]}{\left[B\,J_2(2\sqrt{Z}) + C\,I_2(2\sqrt{Z})\right]} \tag{5.3}$$

Comparison to (2.15) where summations have been carried out to include all frequencies, leads to the same general conclusions as those reached in Section II and represented by equation (2.17).

Thus we conclude that for the double conical bar, the central question of this effort is answered in the affirmative.

SECTION VI

Conclusions and Recommendations

The viability of analytically determining the slope of a simple free-free beam when the displacement and frequency information is known has been demonstrated. The method used was applied to a uniform beam and extended to a beam of double conical circular section. The general equation for a second order, skewed stiffness and mass distribution problem was derived but not solved.

The solution of this last case, probably with the aid of a digital or analog computer, would complete this study and should result in the same conclusion reached above, that the solution to the problem as posed, does in fact exist.

REFERENCES

1. Bisplinghoff, R. L., Ashley, H., Halfman, R. L., "Aeroelasticity," Addison-Wesley Publishing Co., Inc., 1955

2. Timoshenko, S., Young, D. H., Weaver, W. Jr., "Vibration Problems in Engineering," John Wiley and Sons, Inc, 1974

3. Crandall, S. H., Mark, W. D., "Random Vibration in Mechanical Systems," Academic Press, 1963

4. Robson, J. D. "An Introduction to Random Vibration," Edinburgh Univ. Press, 1963

5. Kreyszig, E., "Advanced Engineering Mathematics," John Wiley and Sons, Inc., 1962

6. Thomson, W. T., Barton, M. V., "The Response of Mechanical Systems to Random Excitation," Journal of Applied Mechanics, June, 1957, pages 248-251

7. Nicholson, J. W., "The Lateral Vibrations of Bars of Variable Section," Proceedings at the Royal Society at London, Series A, Vol. 93, 1916-17, pages 506-519

APPENDIX A

Equation of Motion Derived

The equation of motion in beam bending is well known and is derived in many texts on vibrations and aeroelasticity. The method of derivation used here is substantially that followed in Ref. 1, pages 67 to 69.

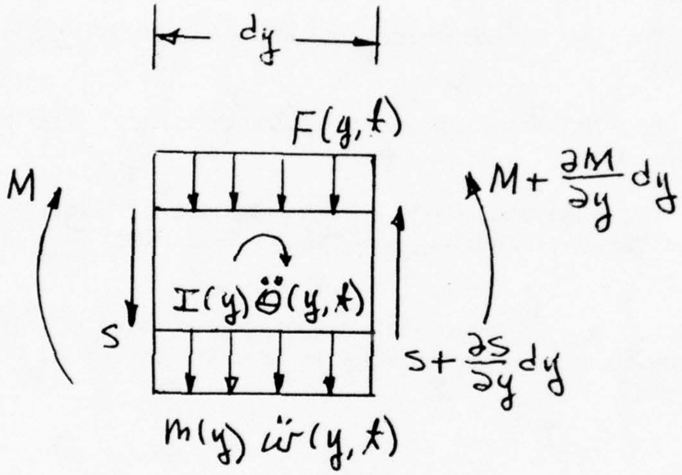

dy
F(y,t)
M
$M+\frac{\partial M}{\partial y}dy$
$I(y)\ddot{\Theta}(y,t)$
S
$S+\frac{\partial S}{\partial y}dy$
$m(y)\ddot{w}(y,t)$

Figure 2. Beam element.

From equilibrium of forces we have

$$m(y)\ddot{w}(y,t)+F(y,t)+S-\left(S+\frac{\partial S}{\partial y}dy\right)$$

or (A.1)

$$m(y)\ddot{w}(y,t)-\frac{\partial S}{\partial y}-F(y,t)=0$$

From equilibrium of moments we have

$$M-S\frac{dy}{2}+I(y)\ddot{\Theta}(y,t)-\left(S+\frac{\partial S}{\partial y}dy\right)\frac{dy}{2}-\left(M+\frac{\partial M}{\partial y}dy\right)=0$$

which becomes

$$\frac{\partial M}{\partial y} + S - I(y)\,\ddot{\theta}(y,t) = 0 \tag{A.2}$$

where higher orders of small quantities have been ignored.

From elementary strength of materials we have the relations

$$\theta'' = \frac{M}{EI} \qquad \text{for bending deflection}$$

and also

$$\phi' = \frac{S}{GK} \qquad \text{for shear deflection.}$$

The total beam deflection is the sum of the shear and bending deflections so that we have

$$w(y,t) = \theta(y,t) + \phi(y,t) \tag{A.3}$$

Combining all these equations gives, after some manipulation

$$m\ddot{w} - \left(\frac{EIm}{GK} + I\right)\ddot{w}'' + \frac{Im}{GK}\ddddot{w} + EI\,w^{IV} = 0 \tag{A.4}$$

Equation (A.4) is practically unsolvable. The more common equation, and the one used in this paper is derived from (A.1) through (A.3) with

G = ∞ and $I(y) = F(y,t) = 0.$

This gives

$$(EIw'')'' + m\ddot{w} = 0 \tag{A.5}$$

Setting G = ∞ implies the beam is infinitely rigid in torsion while setting I(y) = 0 implies that no external loading acts on the beam other than the inertia load m(y) $\ddot{w}$.

APPENDIX B

Stiffness Variation Curve Derived

Referring to Figure 1 and assuming

$$z = Ay^2 + Gy + B \tag{B.1}$$

we have

$$z(-\tfrac{l}{2}) = a = z(\tfrac{l}{2}) \quad ; \qquad z(c) = b \tag{B.2}$$

which gives on substution, the equations

$$\begin{aligned} a &= A\frac{l^2}{4} - G\frac{l}{2} + B \\ a &= A\frac{l^2}{4} + G\frac{l}{2} + B \\ b &= Ac^2 + Gc + B \end{aligned} \tag{B.3}$$

Solving equations (B.3) gives eventually

$$\begin{aligned} A &= \frac{b-a}{c^2 - \frac{1}{4}l^2} \\ B &= a - \frac{l^2}{4}\left(\frac{b-a}{c^2 - \frac{1}{4}l^2}\right) \\ G &= 0 \end{aligned} \tag{B.4}$$

Finally, substituting (B.4) into (B.1) gives the desired equation

$$z = \left[\frac{b-a}{c^2 - \frac{1}{4}l^2}\right] y^2 + a - \frac{l^2}{4}\left[\frac{b-a}{c^2 - \frac{1}{4}l^2}\right] \tag{B.5}$$

which can be written as

$$z = Ay^2 + B \qquad \text{(B.6)}$$

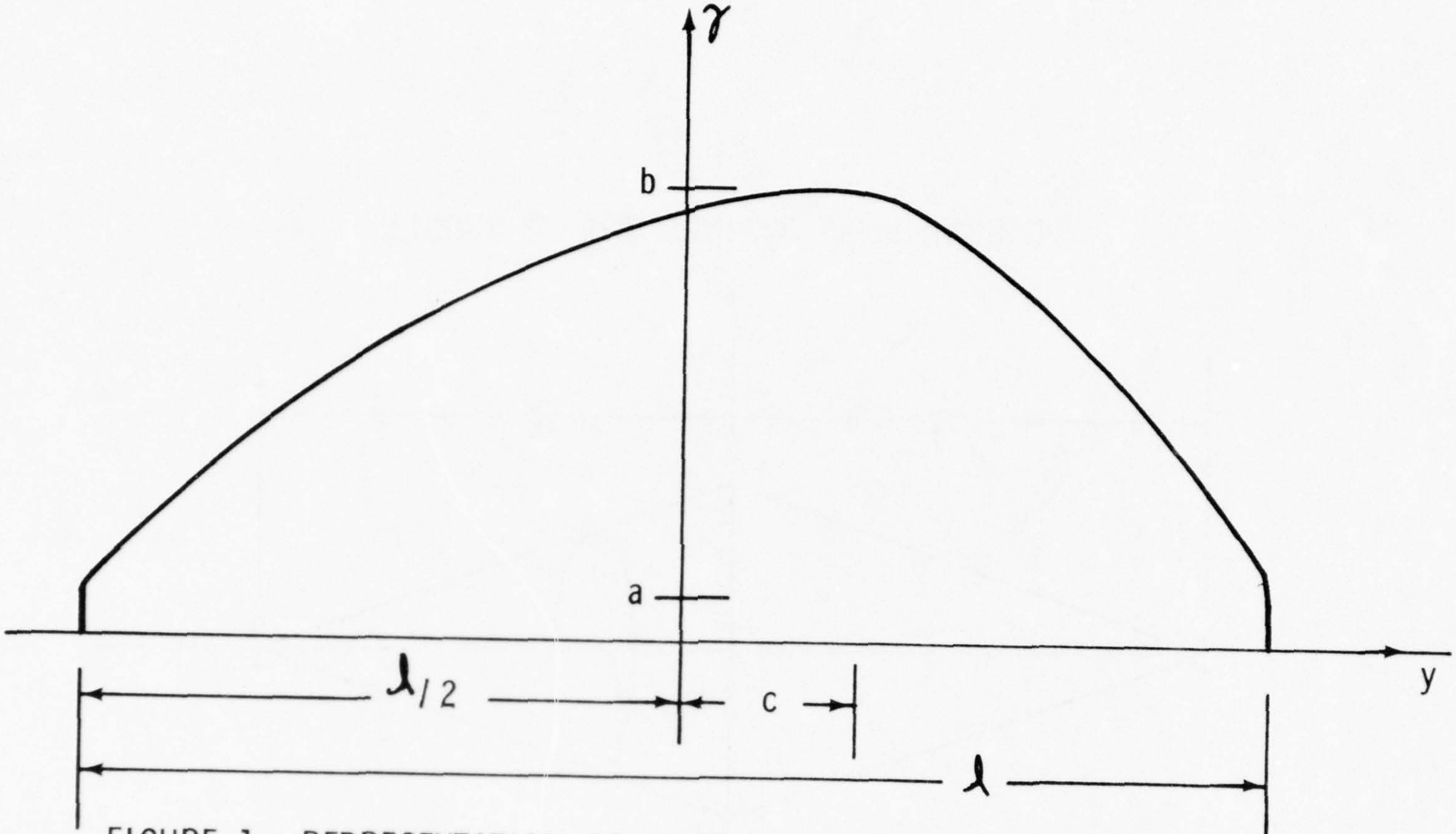


FIGURE 1. REPRESENTATION OF STIFFNESS AND MASS DISTRIBUTION

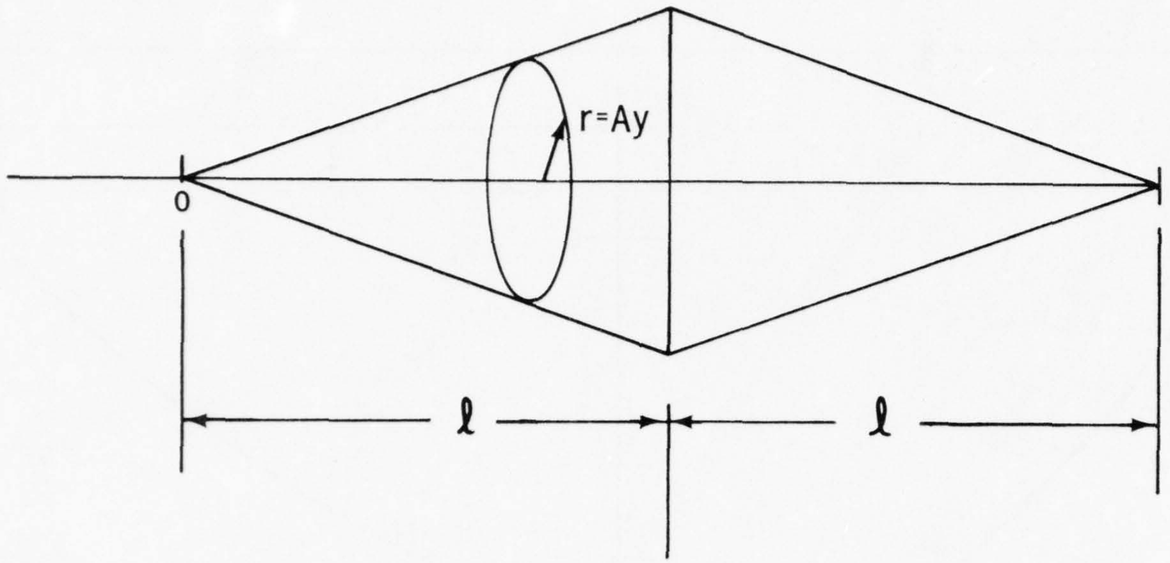


FIGURE 2. THE CONICAL BEAM PROBLEM

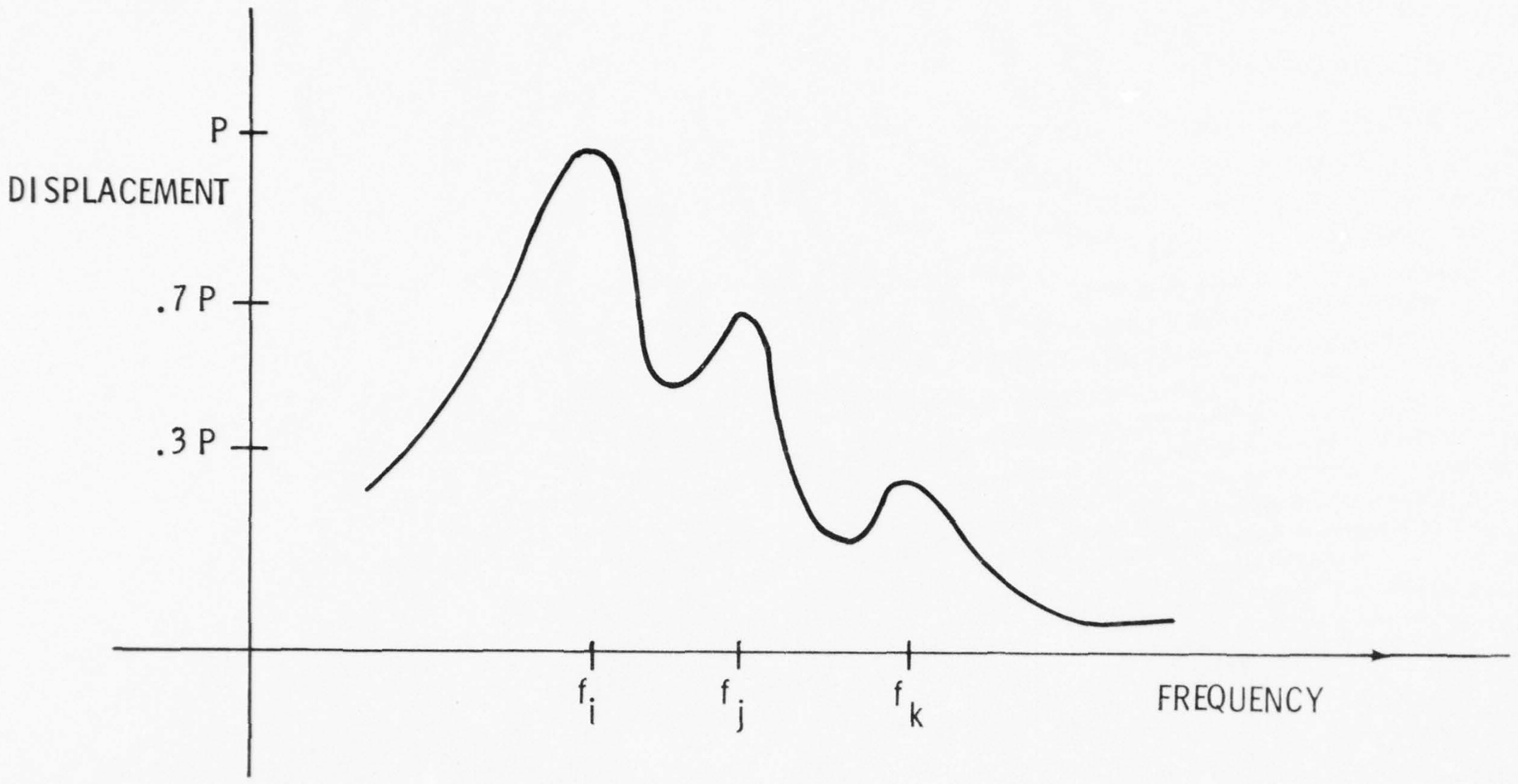


FIGURE 3. DISPLACEMENT RESPONSE VS FREQUENCY

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

ANALYSIS OF INHERENT ERRORS IN ASYNCHRONOUS REDUNDANT DIGITAL FLIGHT CONTROLS

Prepared by:	Charles Slivinsky, PhD.
Academic Rank:	Associate Professor
Department and University:	Department of Electrical Engineering University of Missouri - Columbia
Assignment:	
(Laboratory)	Flight Dynamics
(Division)	Flight Control
(Branch)	Control Systems Development
USAF Research Colleague:	Captain Vincent J. Darcy
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

ANALYSIS OF INHERENT ERRORS IN ASYNCHRONOUS REDUNDANT DIGITAL FLIGHT CONTROLS

by

Charles Slivinsky

ABSTRACT

This report describes the development and application of methods of analyzing inherent errors in asynchronous redundant digital flight controls. Such errors arise because of differences in redundant channel outputs due to the fact that the system sampling rate differs slightly for each channel because of their separately clocked operation. Knowledge of these errors is essential if the monitor for the redundant channels is to be able to distinguish between a failed channel and one operating normally.

The concept of skewed sampling is developed and a single-rate, closed-loop state variable model for realistic aircraft control loops is developed. Using this model a covariance analysis of the channel differences is given. This statistical treatment is based on filtered white noise external inputs, but it is shown how such inputs can be used to generate signals which approximate the true system external inputs. The model and the analysis are general and applicable to a variety of systems. A user-oriented software package, using FORTRAN, is developed to facilitate the required computations and to allow parametric studies of the effect on inherent errors of control system gains, time constants, sample time and other parameters. The analysis is applied to several examples, including the Air Force A7D pitch axis flight control system.

The work provides the basic foundation for further analytical studies of inherent errors. Suggestions are made for refining the model and covariance analysis, to develop additional software and to perform parametric studies. Analytical results can be correlated with Air Force Flight Dynamics Laboratory/ Digital Avionics Integration System (AFFDL/DAIS) laboratory hardware simulations.

ACKNOWLEDGEMENT

It is a pleasure to acknowledge the extensive assistance provided by the author's Research Colleague, Captain Vincent J. Darcy, AFFDL/DAIS Technical Manager. This work is a joint effort of both of us.

SECTION I

INTRODUCTION

Current and projected flight control systems for military aircraft utilize redundancy to achieve the required system reliability. This report describes and analyzes effects which are inherent in redundant, asynchronous digital flight control channels. The term asynchronous is used because the redundant channels are not coordinated in time with each other by any hardware or software linkages. The channels perform identical types of control computations using the same physical variables; however, they are separately and independently clocked.

When compared with synchronized channel operation, asynchronous operation has the advantage of greater reliability due to its less complex hardware structure. It is also less expensive to build and has simpler software requirements.

There are two separate effects in asynchronous operation which must be distinguished. First, channel failures may occur, resulting in improper operation of the failed channels and erroneous channel outputs. Second, under normal, failure-free operation the asynchronous channels will not produce identical results because they are operating on physical variables which are sampled at slightly different times. The sample times are different because the processor crystals, which control the sampling, produce frequencies which are close to each other but are not identical. This second effect gives rise to what is designated as inherent errors between any two redundant channels.

Digital actuator-command voter/monitors are used in the redundant flight control system for each controller axis in order to detect channel failures, isolate them, and to vote; i.e., use the redundant channel outputs to produce a "best" channel output to drive the appropriate aircraft actuator system. The monitor function is accomplished by sets of comparison monitors which compare the actuator-command outputs from all pair-combinations of the redundant channels. Each comparison monitor must allow some differences in numerical values between the channel outputs because asynchronous operation produces the inherent errors mentioned above. Thus, it is important to know the expected magnitudes of inherent errors if these are to be distinguished from channel failures. Therefore, this report considers only inherent errors in order to determine their magnitude in the absence of channel failures.

In Section 2 below asynchronous sampling and voting are discussed. A simple, but adequate approximation to asynchronous sampling is given and a specific voter algorithm is selected and related to a dynamic flight condition. The next section develops a dynamic model of a closed-loop sampled-data flight control system which utilizes two redundant, asynchronous digital control channels and voting. The model is a vector-matrix formulation of linear difference equations describing the system dynamic behavior and the inherent error between the outputs of the two redundant channels.

Section 4 contains a covariance analysis based on the model of Section 3. Expressions are developed for the covariance of the closed-loop system states and for the inherent channel error. Then, two simplified examples are worked out and discussed. They illustrate the application of both the modeling and the covariance analysis.

Section 5 describes the software which has been developed to apply the methods to realistic examples. A general description, a flowchart, and user instructions are given for the FORTRAN program. The next Section contains an application to the A-7D pitch-axis flight control system. Section 7 contains the conclusions and suggestions for future work.

SECTION II

MODELING ASYNCHRONOUS SAMPLING AND VOTING

1. MODEL FOR ASYNCHRONOUS SAMPLING

Asynchronous operation of dual redundant channels can be depicted as shown in Figure 1. A flight control variable is measured by the flight control sensor on the left. This variable is sampled by two asynchronous but identical channels. Each channel processes the variable and produces an output which is sensed by the voter/monitor. The latter computes a single output which is sent to an actuator. The sample periods T_1 and T_2 would be identical in the ideal case. However, these times are computed independently by the separate processors having separate crystal-controlled timing circuitry. Thus T_1 and T_2 will differ by the same percentage that the crystals differ from each other.

An exaggerated description of the effect of separately clocked sampling is shown in Figure 2. The processors start at the same time and after four sample times each of duration T_1, they are back to sampling at the same time. The time required to return to synchronism can be computed as follows.

Every T_1 seconds processor 2 gains a time increment of $T_1 - T_2$ seconds. When the accumulated gain is T_2 seconds the two channels are back in synchronism. Let P be equal to the period required to accumulate T_2 seconds. Then in P seconds, T_2 seconds have been gained, or

$$\frac{P}{T_1} \times (T_1 - T_2) = T_2 \qquad (1)$$

or

$$P = \frac{T_1 T_2}{T_1 - T_2} \qquad (2)$$

Let e be the percentage error between the crystals controlling the separate processors; i.e.,

$$e = \frac{T_1 - T_2}{T_1} \times 100 \qquad (3)$$

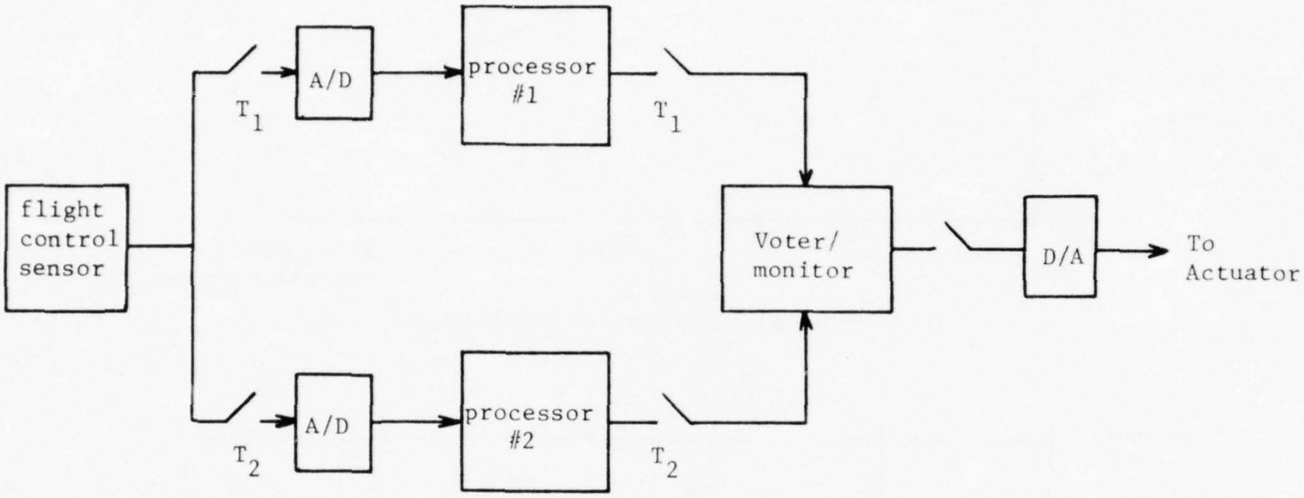


Figure 1.
Asynchronous, Dual-Redundant System

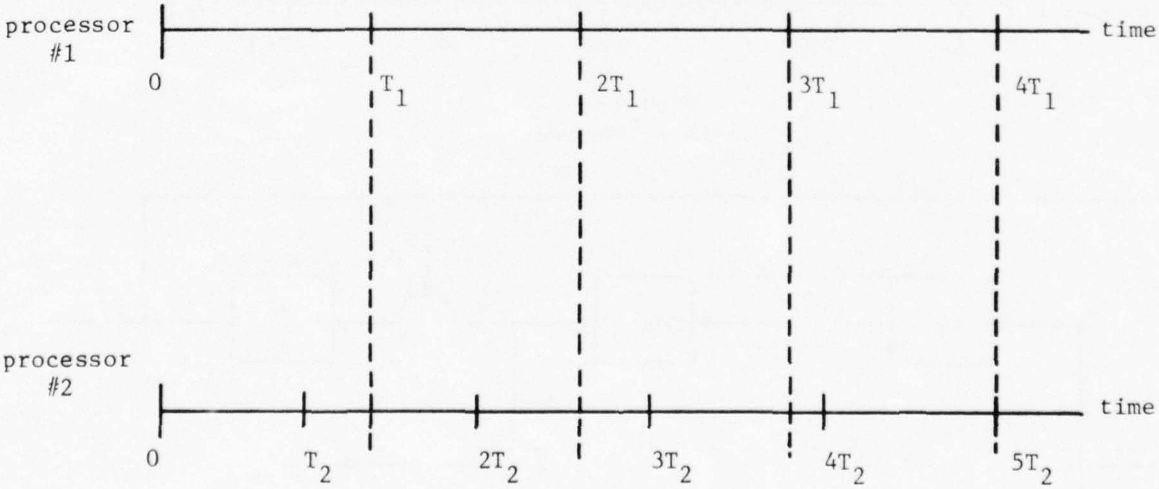


Figure 2.
Asynchronous, Separately Clocked Operation

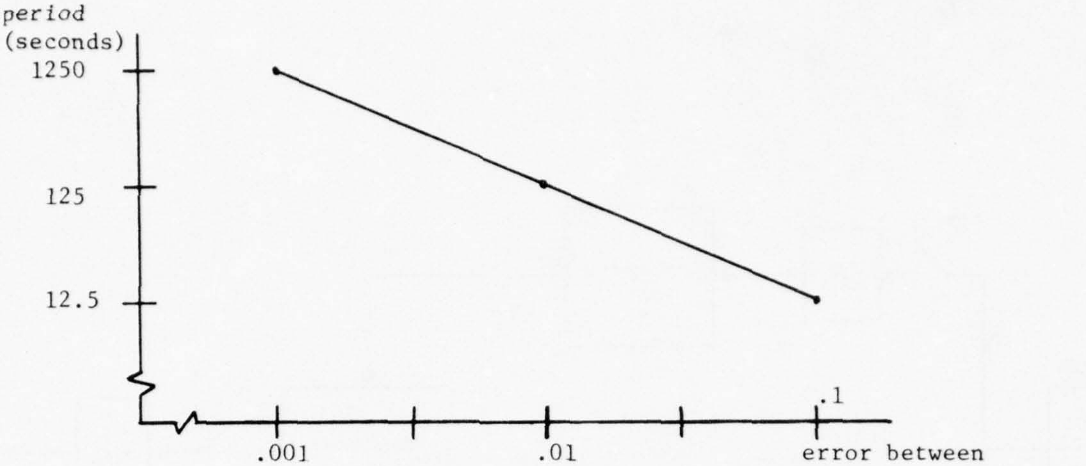


Figure 3.
Time-To-Synchronization vs. Clock Error

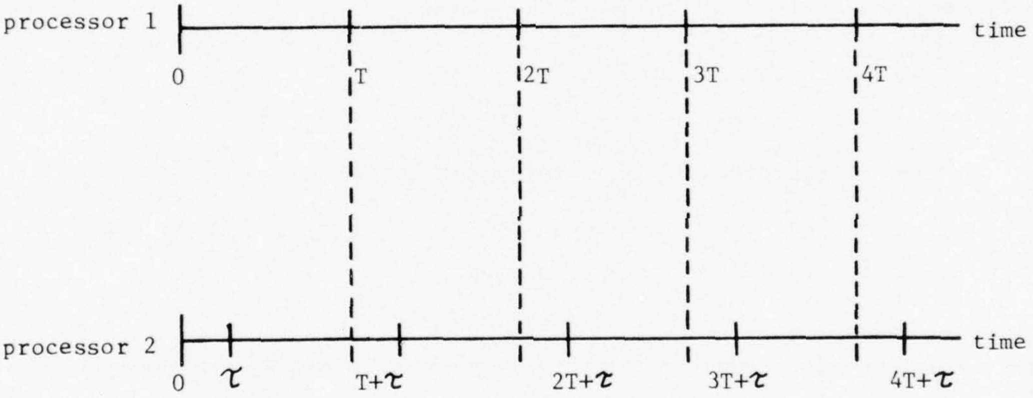


Figure 4.
Skewed Sampling

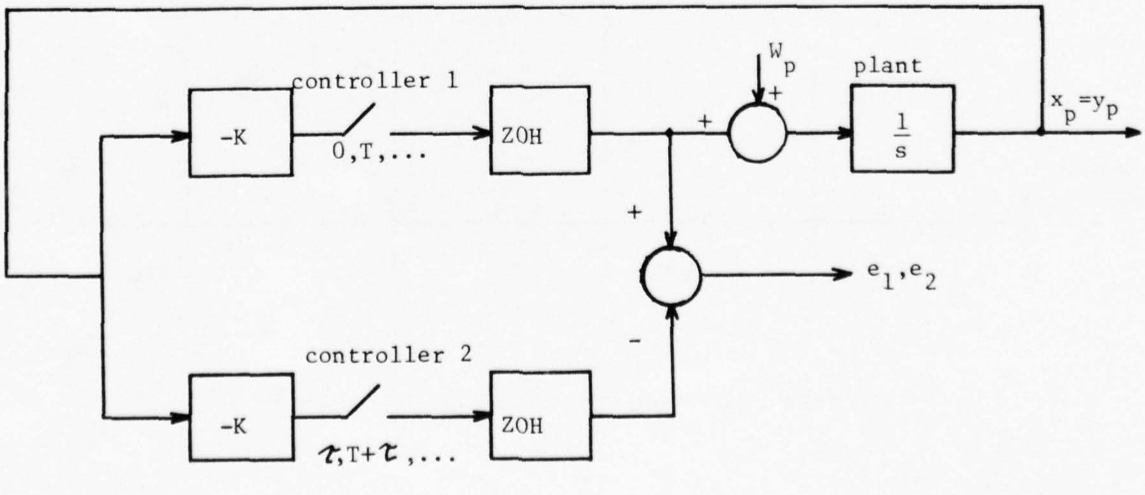


Figure 5.
Example 1

In terms of e the period P is

$$P = \frac{100 T_1}{e} - T_1 \approx \frac{100 T_1}{e} \tag{4}$$

A logarithmic plot of P versus e for T_1 = 1/80 seconds is shown in Figure 3.

For realistic values of e the plot shows that P is several minutes or more. That is, the relative positions of the sample times of the two channels will change very slowly. With an e of 0.01%, 10,000 samples are required to regain synchronism. Thus over a shorter time interval it is plausible to consider the relative position to be fixed. This means that the two channels are operating at exactly the same rate but there is a fixed offset or time skew between them.

Skewed sampling is shown in Figure 4. The parameter τ is the amount of skew. Skewed sampling will be used from this point on to simplify the modeling process.

2. MODEL FOR VOTING

Various algorithms for the operation of the voter portion of the asynchronous voter/monitor are available. Three are discussed briefly here.

The first scheme uses the output of the most recently updated channel as its output. Its advantage is that the effective sample time of the sampled data system is reduced because the control signals are sent at a rate which is always greater than or equal to 1/T. This reduces the lag effect due to sampling and results in better dynamic performance. Its disadvantage is that no selection is made from among the channels and this could possibly result in sending out an output from a failed channel during times when the monitoring function fails to detect the failure for one reason or another.

The next scheme makes use of averaging of the channel outputs. Its advantage is that the adverse effects of sensor noise are also averaged and, as a result, are probably reduced. Its disadvantage is that the most recent data is averaged along with "older" data and so the net sample rate is longer than that for the previous scheme.

The third scheme is labelled "median select". It requires that the channel outputs be compared with each other and that the upper (for a four-channel system) or lower (for a three-or four-channel system) median be used as the voter output. It has the advantage of inherently discarding a comparatively large or small single channel output, but does not give as high as effective sample rate as the first scheme.

In the modeling below the voting scheme will be to select the same channel output at all times. This scheme is roughly equivalent to the third scheme above when the channel outputs are either monotonically increasing or decreasing in time.

SECTION III

STATE EQUATIONS FOR CLOSED-LOOP OPERATION

1. PLANT EQUATIONS

The overall closed-loop dynamic system consists of a continuous-time plant and dual-redundant single-rate discrete time controllers. The plant output is sampled by each of the controllers, using a common sample period but having a fixed skew between them. The output of one of the controllers serves as the piecewise-constant input to the plant, along with an external input vector.

The plant equations include aircraft, sensor and actuator dynamics, as well as any dynamics associated with the pilot input and wind-gust model input. They are assumed to be in the form

$$\dot{x}_p = A_p x_p + B_{1p} u_p + B_{2p} w_p \qquad (5)$$

$$y_p = C_p x_p \qquad (6)$$

where

x_p = plant state vector $(n_p \times 1)$
u_p = plant input vector $(n_{up} \times 1)$
w_p = disturbance input vector $(n_{wp} \times 1)$
y_p = plant output vector $(n_{op} \times 1)$
A_p = plant state matrix $(n_p \times n_p)$
B_{1p} = plant control input matrix $(n_p \times n_{up})$
B_{2p} = plant disturbance input matrix $(n_p \times n_{wp})$
C_p = plant output matrix $(n_{op} \times n_p)$

The solution to equation (5) is

$$x_p(t) = \Phi(t,t_o)\, x_p(t_o) + \int_{t_o}^{t} \Phi(t,s) B_{1p}\, u_p(s)\, ds + \int_{t_o}^{t} \Phi(t,s)\, B_{2p}\, w_p(s)\, ds \qquad (7)$$

where $\Phi(t,t_o)$ is the state transition matrix and for constant A_p is given by

$$\Phi(t,t_o) = \exp\left[A_p(t-t_o)\right] \qquad (8)$$

The plant input $u_p(t)$ is piecewise constant over a given sampling interval; i.e.,

$$u_p(t) = u_p(t_k) \qquad t_k \leq t < t_{k+1} \qquad (9)$$

and so for $t_o = t_k$, $t = t_{k+1}$, and $t_{k+1} - t_k = T$, the second term in equation (7) can be written as

$$\int_{t_k}^{t_{k+1}} \Phi(t_{k+1},s)\, B_{1p}\, u_p(s)\, ds = \psi(t_{k+1},t_k)\, u_p(t_k) \tag{10}$$

where

$$\psi(t_{k+1},t_k) = \int_0^T \exp(A_p t)\, B_{1p}\, dt \tag{11}$$

Substitution of (10) into (7) gives

$$\begin{aligned} x_p(t_{k+1}) = \Phi(t_{k+1},t_k)\, x_p(t_k) &+ \psi(t_{k+1},t_k)\, u_p(t_k) \\ &+ \int_{t_k}^{t_{k+1}} \Phi(t_{k+1},s)\, B_{2p}\, w_p(s)\, ds \end{aligned} \tag{12}$$

2. CONTROLLER EQUATIONS

There are two asynchronous redundant controllers. The first satisfies the following vector difference equations

$$x_{c1}(t_{k+1}) = F_c\, x_{c1}(t_k) + G_c\, u_{c1}(t_k) \tag{13}$$

$$y_{c1}(t_k) = H_c\, x_{c1}(t_k) + E_c\, u_{c1}(t_k) \tag{14}$$

and the second satisfies

$$x_{c2}(t_{k+1}+\tau) = F_c\, x_{c2}(t_k+\tau) + G_c\, u_{c2}(t_k+\tau) \tag{15}$$

$$y_{c2}(t_k+\tau) = H_c\, x_{c2}(t_k+\tau) + E_c\, u_{c2}(t_k+\tau) \tag{16}$$

where

x_{c1} = controller 1 state vector $(n_c \times 1)$
y_{c1} = controller 1 output vector $(m_{up} \times 1)$
x_{c2} = controller 2 state vector $(n_c \times 1)$
y_{c2} = controller 2 output vector $(m_{up} \times 1)$
F_c = controller state matrix $(n_c \times n_c)$
G_c = controller control input matrix $(n_c \times m_{op})$
H_c = controller output output matrix (states) $(m_{up} \times n_c)$
E_c = controller output matrix (inputs) $(m_{up} \times m_{op})$

The matrices F_c, G_c, H_c, and E_c could be functions of the sample time T.

3. TOTAL TRANSITION EQUATIONS

The plant equations and the controller equations are related by the requirements that the control input to the plant is the output of controller 1 and the plant output is the input to each controller; i.e.,

$$u_p(t_k) = y_{c1}(t_k) \tag{17}$$

$$u_{c1}(t_k) = y_p(t_k) \tag{18}$$

$$u_{c2}(t_k+\tau) = y_p(t_k+\tau) \tag{19}$$

Substitution of equations (6), (14) and (18) into (17) gives the plant input in terms of the plant and controller 1 state variables, as

$$u_p(t_k) = H_c x_{c1}(t_k) + E_c C_p x_p(t_k) \tag{20}$$

and substitution of (20) into the plant state equation (12) gives $x_p(t_{k+1})$ in terms of $x_p(t_k)$, $x_{c1}(t_k)$, and w_p as

$$\begin{aligned} x_p(t_{k+1}) &= \left[\Phi(t_{k+1},t_k) + \psi(t_{k+1},t_k)E_cC_p\right]x_p(t_k) \\ &\quad + \psi(t_{k+1},t_k)H_c x_{c1}(t_k) \\ &\quad + \int_{t_k}^{t_{k+1}} \Phi(t_{k+1},s)B_{2p}w_p(s)ds \end{aligned} \tag{21}$$

In a similar manner, substituting (6) and (19) into the controller 1 equations (13) and (14) gives

$$x_{c1}(t_{k+1}) = F_c x_{c1}(t_k) + G_c C_p x_p(t_k) \tag{22}$$

$$y_{c1}(t_k) = H_c x_{c1}(t_k) + E_c C_p x_p(t_k) \tag{23}$$

For controller 2, equations (6) and (19) show that

$$u_{c2}(t_k+\tau) = C_p x_p(t_k+\tau) \tag{24}$$

The quantity $x_p(t_k+\tau)$ can be written using the solution to equation (7) as

$$x_p(t_k+\tau) = \left[\Phi(t_k+\tau, t_k) + \Psi(t_k+\tau, t_k) E_c C_p\right] x_p(t_k)$$
$$+ \Psi(t_k+\tau, t_k) H_c x_{c1}(t_k)$$
$$+ \int_{t_k}^{t_k+\tau} \Phi(t_k+\tau, s) B_{2p} w_p(s) ds \tag{25}$$

By substituting (24) and (25) into (15) and (16) the controller 2 equations are obtained as

$$x_{c2}(t_{k+1}+\tau) = F_c x_{c2}(t_k+\tau)$$
$$+ G_c C_p \left[\Phi(t_k+\tau, t_k) + \Psi(t_k+\tau, t_k) E_c C_p\right] x_p(t_k)$$
$$+ G_c C_p \Psi(t_k+\tau, t_k) H_c x_{c1}(t_k)$$
$$+ G_c C_p \int_{t_k}^{t_k+\tau} \Phi(t_k+\tau, s) B_{2p} w_p(s) ds \tag{26}$$

$$y_{c2}(t_k+\tau) = H_c x_{c2}(t_k+\tau)$$
$$+ E_c C_p \left[\Phi(t_k+\tau, t_k) + \Psi(t_k+\tau, t_k) E_c C_p\right] x_p(t_k)$$
$$+ E_c C_p \Psi(t_k+\tau, t_k) H_c x_{c1}(t_k) + E_c C_p \int_{t_k}^{t_k+\tau} \Phi(t_k+\tau, s) B_{2p} w_p(s) ds \tag{27}$$

Equations (12), (21), and (26) describe the closed-loop system in state variable form as a set of linear difference equations. One additional quantity is needed, the difference between the two controller outputs. This quantity measures the inherent error due to asynchronous operation. There are two ways to define the error. In the interval $t_k + \tau < t \leq t_{k+1}$,

$$e_1(t) = y_{c1}(t_k) - y_{c2}(t_k+\tau) \tag{28}$$

and for $t_{k+1} < t \leq t_{k+1} + \tau$,

$$e_2(t) = y_{c1}(t_{k+1}) - y_{c2}(t_k+\tau) \tag{29}$$

These equations can be put in compact form by writing them in terms of a combined state vector

$$x(t_k) = \begin{bmatrix} x_p(t_k) \\ x_{c1}(t_k) \\ x_{c2}(t_k+\tau) \end{bmatrix}, \quad x(t_{k+1}) = \begin{bmatrix} x_p(t_{k+1}) \\ x_{c1}(t_{k+1}) \\ x_{c2}(t_{k+1}+\tau) \end{bmatrix} \tag{30}$$

The state equations become

$$x(t_{k+1}) = F(T,\tau)\, x(t_k) + \begin{bmatrix} \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, s) B_{2p} w_p(s)\, ds \\ 0 \\ G_c C_p \int_{t_k}^{t_k+\tau} \Phi(t_k+\tau, s) B_{2p} w_p(s)\, ds \end{bmatrix} \tag{31}$$

where

$$F(T,\tau) = \begin{bmatrix} \Phi(t_{k+1}, t_k) + \psi(t_{k+1}, t_k) E_c C_p & \psi(t_{k+1}, t_k) H_c & 0 \\ G_c C_p & F_c & 0 \\ G_c C_p[\Phi(t_k+\tau, t_k) + \psi(t_k+\tau, t_k) E_c C_p] & G_c C_p \psi(t_k+\tau, t_k) & F_c \end{bmatrix} \tag{32}$$

The controller output equations are

$$y_{c1}(t_k) = H_1 x(t_k) \tag{33}$$

$$y_{c2}(t_k+\tau) = H_2 x(t_k) + E_c C_p \int_{t_k}^{t_k+\tau} \Phi(t_k+\tau, s) B_{2p} w_p(s)\, ds \tag{34}$$

where

$$H_1 = \begin{bmatrix} E_c C_p & H_c & 0 \end{bmatrix} \tag{35}$$

$$H_2 = \begin{bmatrix} E_c C_p[\Phi(t_k+\tau, t_k) + \psi(t_k+\tau, t_k) E_c C_p] & E_c C_p \psi(t_k+\tau, t_k) H_c & H_c \end{bmatrix} \tag{36}$$

The expressions for $e_1(t)$ and $e_2(t)$ become

$$e_1(t) = (H_1 - H_2)\, x(t_k) - E_c C_p \int_{t_k}^{t_k+\tau} \Phi(t_k+\tau, s)\, B_{2p}\, w_p(s)\, ds \tag{37}$$

$$\begin{aligned} e_2(t) = (H_1 F - H_2)\, x(t_k) &+ \begin{bmatrix} H_1 \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, s)\, B_{2p}\, w_p(s)\, ds \\ 0 \\ H_1 G_c C_p \int_{t_k}^{t_k+\tau} \Phi(t_k+\tau, s)\, B_{2p}\, w_p(s)\, ds \end{bmatrix} \\ &- E_c C_p \int_{t_k}^{t_k+\tau} \Phi(t_k+\tau, s)\, B_{2p}\, w_p(s)\, ds \end{aligned} \tag{38}$$

This completes the description of the closed-loop system.

SECTION IV

COVARIANCE ANALYSIS AND EXAMPLES

1. ASSUMPTIONS IN THE ANALYSIS

The covariance analysis described in this section uses the model of Section III. The external input w_p is assumed to be a Gaussian white noise random process with zero mean. In a typical flight control application this input could be used to simulate the spectrum of frequencies of the actual pilot input by first passing the white noise signal through an appropriate linear filter. The filter is chosen so that the frequency spectral density of its output is representative of that of a typical pilot input. The dynamic equations of the filter are incorporated into the plant equations. (See Section VI for an example of this approach.)

Since the mean of the input is zero, the mean of the state $x(t_k)$ is also zero, but the covariance of $x(t_k)$ will be non-zero. A difference equation whose solution gives this covariance is developed below. Also given are expressions for the covariances of the two inherent channel errors $e_1(t)$ and $e_2(t)$. The covariances of $e_1(t)$ and $e_2(t)$ are the measures of the magnitude of the inherent error. Two low-order examples illustrate the application of the results.

2. COVARIANCE OF THE STATES

The state equation for the closed-loop system is given by equation (31), repeated as

$$x(t_{k+1}) = F(T,\tau)\, x(t_k) + \begin{bmatrix} \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, s)\, B_{2p}\, w_p(s)\, ds \\ 0 \\ G_c C_p \int_{t_k}^{t_k+\tau} \Phi(t_k+\tau, s)\, B_{2p}\, w_p(s)\, ds \end{bmatrix} \tag{39}$$

Let the input $w_p(t)$ be a Gaussian white noise random process with zero mean which is independent of $x(0)$.[1] Let $E\{\cdot\}$ indicate the expected value; then,

$$E\{w_p(t_k)\} = 0 \tag{40}$$

$$E\{w_p(t_k)\,x^T(0)\} = 0 \tag{41}$$

$$E\{w_p(t_k)\,w_p^T(t_m)\} = W\,\delta(t_k - t_m) \tag{42}$$

where W is the input disturbance covariance matrix and $\delta(t)$ is the Dirac delta function.

The quantity being sought is $P_x(k)$, the covariance matrix of the states. $P_x(k)$ is defined as

$$P_x(k) = E\{x(t_k)\,x^T(t_k)\} \tag{43}$$

A difference equation for $P_x(k)$ will be developed. $P_x(k+1)$ is given by

$$P_x(k+1) = E\{x(t_{k+1})\,x^T(t_{k+1})\} \tag{44}$$

The quantity $x(t_{k+1})\,x^T(t_{k+1})$ can be obtained by using (3) and carrying out the required multiplications, as

$$\begin{aligned}
x(t_{k+1})\,x^T(t_{k+1}) &= F(T,\tau)\,x(t_k)\,x^T(t_k)\,F^T(T,\tau) \\
&+ F\left[\int_{t_k}^{t_{k+1}} \underbrace{\Phi(t_{k+1},s)\,B_{2p}\,w_p(s)\,ds}_{\alpha} \quad 0 \quad G_c C_p \int_{t_k}^{t_k+\tau} \underbrace{\Phi(t_k+\tau,s)\,B_{2p}\,w_p(s)}_{\beta}\,ds\right] \\
&+ \begin{bmatrix} \int_{t_k}^{t_{k+1}} \Phi(t_{k+1},s)\,B_{2p}\,w_p(s)\,ds \\ 0 \\ G_c C_p \int_{t_k}^{t_k+\tau} \Phi(t_k+\tau,s)\,B_{2p}\,w_p(s)\,ds \end{bmatrix} \\
&+ \begin{bmatrix} \int_{t_k}^{t_{k+1}} \Phi(t_{k+1},s)\,B_{2p}\,w_p(s)\,ds \int_{t_k}^{t_{k+1}} w_p^T(s)\,B_{2p}^T\,\Phi(t_{k+1},s)\,ds & 0 & \int_{t_k}^{t_{k+1}} \alpha\,ds \int_{t_k}^{t_k+\tau} \beta^T ds \\ 0 & 0 & 0 \\ \int_{t_k}^{t_k+\tau} \beta\,ds \int_{t_k}^{t_{k+1}} \alpha^T ds & 0 & \int_{t_k}^{t_k+\tau} \beta\,ds \int_{t_k}^{t_k+\tau} \beta^T ds \end{bmatrix}
\end{aligned} \tag{45}$$

Next, the expected value of both sides of equation (45) is taken. The first term on the right hand side becomes

$$E\left\{F(T,\tau)\,x(t_k)\,x^T(t_k)\,F^T(T,\tau)\right\} = F(T,\tau)\,P_x(k)\,F^T(T,\tau) \tag{46}$$

The next two terms are zero because the expected value of $w_p(t)$ is zero. For the last term consider the expected value of upper left hand matrix:

$$E\left\{\int_{t_k}^{t_{k+1}} \Phi(t_{k+1},\sigma)\,B_{2p}\,w_p(\sigma)\,d\sigma \int_{t_k}^{t_{k+1}} w_p^T(s)\,B_{2p}^T\,\Phi^T(t_{k+1},s)\,ds\right\} =$$

$$\int_{t_k}^{t_{k+1}}\int_{t_k}^{t_{k+1}} \Phi(t_{k+1},\sigma)\,B_{2p}\,E\left\{w_p(\sigma)\,w_p^T(s)\right\}B_{2p}^T\,\Phi^T(t_{k+1},s)\,ds\,d\sigma \tag{47}$$

Using (42) and performing the inner integration, one obtains for the right hand side

$$\int_{t_k}^{t_{k+1}} \Phi(t_{k+1},s)\,B_{2p}\,W\,B_{2p}^T\,\Phi^T(t_{k+1},s)\,ds$$

Define

$$V_0(t) = \int_0^t \Phi(t,s)\,B_{2p}\,W\,B_{2p}^T\,\Phi^T(t,s)\,ds \tag{48}$$

so that the previous expression becomes $V_0(T)$.

Similar developments for the remaining parts of the last term of (45) enable (44) to be written as

$$P_x(k+1) = F(T,\tau)\,P_x(k)\,F^T(T,\tau) + V(T,\tau) \tag{49}$$

where

$$V(T,\tau) = \begin{bmatrix} V_0(T) & 0 & V_0(\tau)[G_cC_p]^T \\ 0 & 0 & 0 \\ G_cC_pV_0(T) & 0 & G_cC_pV_0(\tau)[G_cC_p]^T \end{bmatrix} \tag{50}$$

Equation (49) is the desired difference equation for the covariance of the states. If only the steady-state covariance, designated P_{xss}, is needed, then the following linear matrix equation must be solved for P_{xss}

$$P_{xss} = F(T,\tau) P_{xss} F^T(T,\tau) + V(T,\tau) \tag{51}$$

3. COVARIANCE OF THE ERRORS

The covariances of $e_1(t)$ and $e_2(t)$ are calculated using the same procedure as in the previous development. Let P_{e1} be the covariance of e_1. Then

$$P_{e1}(k) = E\left\{ e_1(t_k) e_1^T(t_k) \right\} \tag{52}$$

Using equation (37) for $e_1(t)$ and taking expected values results in

$$P_{e1}(k) = (H_1 - H_2) P_x(k) (H_1 - H_2)^T + E_c C_p V_o(\tau) (E_c C_p)^T \tag{53}$$

Thus once $P_x(k)$ is known, $P_{e1}(k)$ can be calculated by performing the required matrix multiplications.

Let P_{e2} be the covariance of e_2. The development of the expression for P_{e2} is tedious and many terms involving $V_o(T)$ and $V_o(\tau)$ are produced which cancel out once the specific structure of H_1 and H_2 (equations (35) and (36)) are incorporated. The result is

$$P_{e2}(k) = \left[H_1 F(T,\tau) - H_2\right] P_x(k) \left[H_1 F(T,\tau) - H_2\right]^T + E_c C_p \left[V_o(T) - V_o(\tau)\right] (E_c C_p)^T \tag{54}$$

The application of state variable modeling and equations (51), (53), and (54) is illustrated next.

4. EXAMPLES

As the first example consider the closed-loop system shown in Figure 5. The plant is described by $A_p = 0$, $B_{1p} = 1$, $B_{2p} = 1$, and $C_p = 1$. The controllers are each described by $F_c = 0$, $G_c = 1$, $H_c = 0$, and $E_c = -K$. Assume a white noise, zero mean input with $W = \sigma_w^2$. The state transition matrix $\Phi(t_2,t_1)$, from (8), is

$$\begin{aligned}\Phi(t_2,t_1) &= \exp[A_p(t_2-t_1)\\ &= 1\end{aligned} \qquad (55)$$

and $\psi(t_2,t_1)$, according to (11), is $t_2 - t_1$.

Using (32) the matrix $F(T,\tau)$ is calculated to be

$$F(T,\tau) = \begin{bmatrix} 1-TK & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad (56)$$

$V_o(t)$, from (48), is $\sigma_w^2 t$, and $V(T,\tau)$, from (50) is

$$V(T,\tau) = \begin{bmatrix} \sigma_w^2 T & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad (57)$$

The steady-state covariance of the states is found by solving (51) for P_{xss}, resulting in

$$P_{xss} = \frac{\sigma_w^2}{K(2-TK)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad (58)$$

P_{elss} and P_{e2ss} can be calculated from (53) and (54). For this example H_1 is [-k 0 0], H_2 is [-k(1- k) 0 -k] and so

$$P_{e1ss} = K^2 \tau \sigma_w^2 \left(1 + \frac{K\tau}{2-TK} \right) \tag{59}$$

$$P_{e2ss} = K^2 (T-\tau) \sigma_w^2 \left(1 + \frac{K(T-\tau)}{2-TK} \right) \tag{60}$$

P_{e1ss} and P_{e2ss} are plotted in Figure 6. The diagrams to the right of each plot show the times at which the controller outputs are sampled for the calculation of e_1 and e_2. The plots show the error variances are periodic since the effective skew between the channels varies periodically between the limits 0 and T. Also, the error variances are largest when the times at which y_{c1} and y_{c2} change are farthest apart, as expected. The results indicate that some combination of the two possible ways of measuring the channel errors may be less effected by the amount of skew than either error taken alone.

As a second example consider the closed-loop system of Figure 7. Here the model parameters are

$$\begin{array}{lll} A_p = 0 & F_c = 0 & W = \sigma_w^2 \\ B_{1p} = 1 & G_c = 1 & \Phi(t_2, t_1) = 1 \\ B_{2p} = 1 & H_c = -K & \psi(t_2, t_1) = t_2 - t_1 \\ C_p = 1 & E_c = 0 & \end{array} \tag{61}$$

so that

$$F(T,\tau) = \begin{bmatrix} 1 & -KT & 0 \\ 1 & 0 & 0 \\ 1 & -K\tau & 0 \end{bmatrix} \tag{62}$$

$$V_0(t) = \sigma_w^2 t \tag{63}$$

$$V(T,\tau) = \sigma_w^2 \begin{bmatrix} T & 0 & \tau \\ 0 & 0 & 0 \\ \tau & 0 & \tau \end{bmatrix} \tag{64}$$

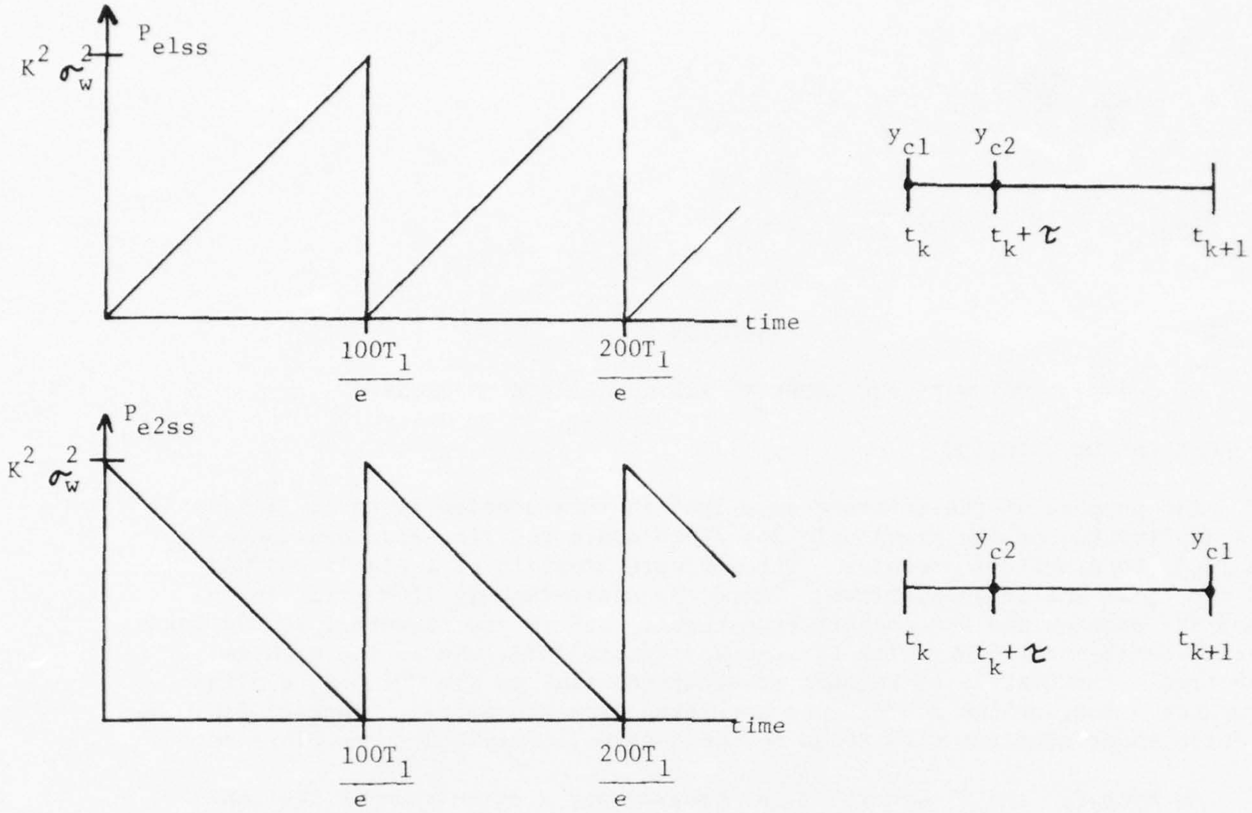


Figure 6.
P_{e1ss} And P_{e2ss} For Example 1

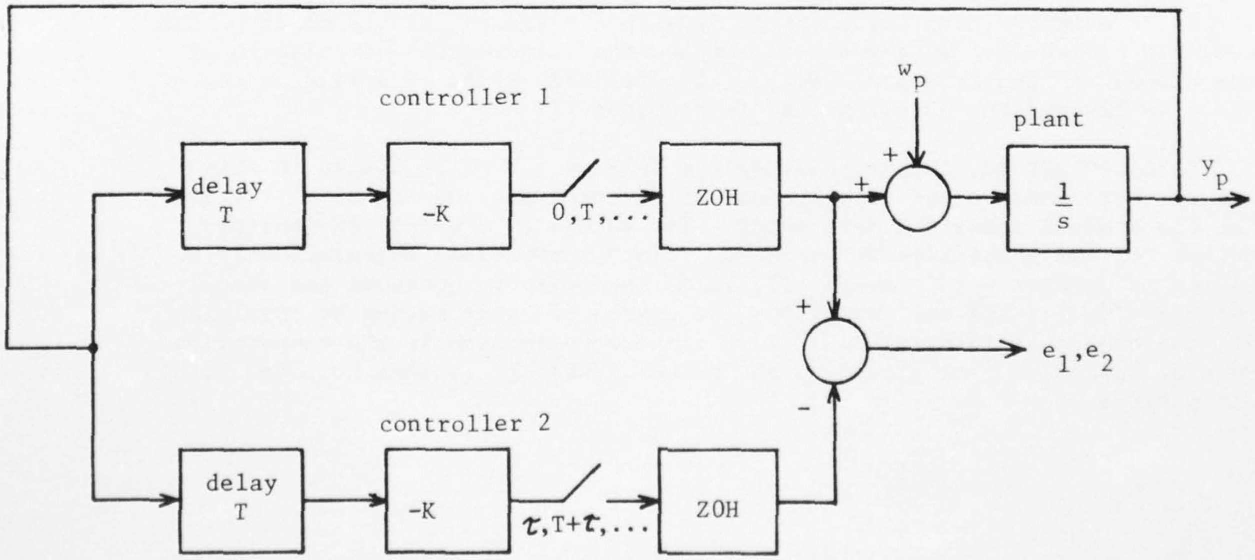


Figure 7.
Example 2

$$\text{and } P_{e1ss} = K^2 \tau \sigma_w^2 \left(1 + \frac{K\tau}{2-TK} \right) \quad (65)$$

$$P_{e2ss} = K^2 (T-\tau) \sigma_w^2 \left(1 + \frac{K(T-\tau)}{2-TK} \right) \quad (66)$$

SECTION V

SOFTWARE FOR MODELING AND COVARIANCE ANALYSIS

1. GENERAL DESCRIPTION

The purpose of the software described in this section is to facilitate the application of the previously described state modeling and covariance analyses to practical examples. The software consists of a single FORTRAN main program and 16 subroutines. There are approximately 1500 cards in the complete package and the core storage requirement is approximately 30,000 words for a fourth-order plant with first-order controllers; the latter problem requires approximately 60 seconds of execution time on the CDC 6600 digital computer. Most of the subroutines are taken from the software package DIGIKON, written under contract with AFFDL by the Honeywell Flight Systems Division.[2]

An overview and flowchart of the program and a discussion of its subroutines are given below, along with instructions on how to use it.

2. FLOWCHART AND DESCRIPTION OF MAJOR COMPONENTS AND SUBROUTINES

The main program is coded in a straight-line manner with no major loops. Its major computational tasks are to develop the state variable model of the complete closed-loop system and to compute the steady-state covariances of the states and the errors e_1 and e_2. The variable names were selected to resemble closely the notation used in Sections III and IV.

A flowchart for the program appears in Figure 8. The blocks in this figure correspond to the clearly identified components of the main program. The first block shows the data input. The variables are self-explanatory except for the quantities NT and NTAU. Since a numerical integration is required to compute $V_o(\tau)$ and $V_o(T)$, it is necessary to quantize the time interval [0,T]. The user specifies the degree of quantization by supplying NT, the number of intervals in [0,T] which are to be used in the computation. He also specifies τ by providing the number NTAU; τ is then computed within the program as

$$\tau = \frac{\text{NTAU}}{\text{NT}} \times T \quad (67)$$

The second block specifies the calculation of $\Phi(\tau)$, $\psi(\tau)$, $\Phi(T)$, and $\psi(T)$. This is accomplished by using the DIGIKON subroutine EXPK2 which is

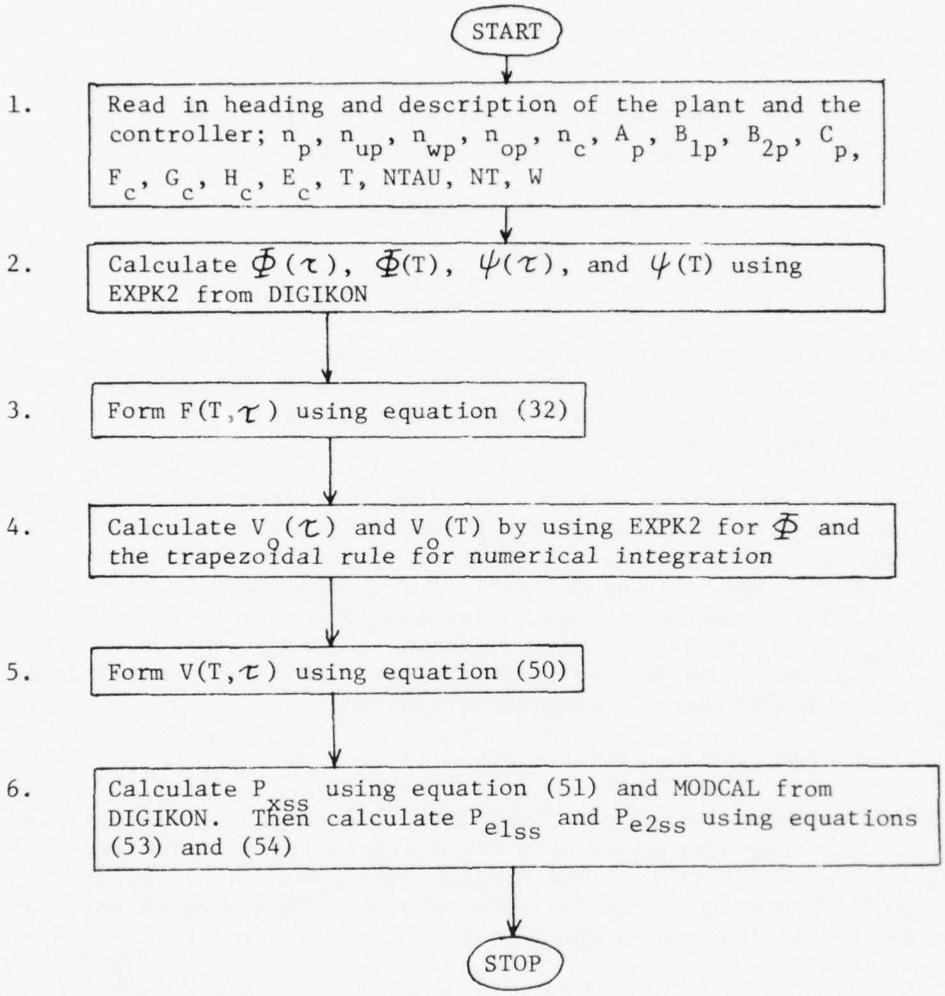

START
1.
Read in heading and description of the plant and the controller; n_p, n_{up}, n_{wp}, n_{op}, n_c, A_p, B_{1p}, B_{2p}, C_p, F_c, G_c, H_c, E_c, T, NTAU, NT, W
2.
Calculate $\Phi(\tau)$, $\Phi(T)$, $\psi(\tau)$, and $\psi(T)$ using EXPK2 from DIGIKON
3.
Form $F(T,\tau)$ using equation (32)
4.
Calculate $V_o(\tau)$ and $V_o(T)$ by using EXPK2 for Φ and the trapezoidal rule for numerical integration
5.
Form $V(T,\tau)$ using equation (50)
6.
Calculate P_{xss} using equation (51) and MODCAL from DIGIKON. Then calculate P_{e1ss} and P_{e2ss} using equations (53) and (54)
STOP

Figure 8.
Flowchart Describing the Major Computations

fully described in the Technical Reports of reference 2. Next, $F(T,\tau)$, the discrete-time state matrix for the overall system is computed and written. This completes the state variable modeling.

Block 4 indicates that the matrices $V_o(\tau)$ and $V_o(T)$ are next computed. These computations require the use of EXPK2 to compute $\Phi(t_2,t_1)$ at the arguments

$$t_1 = 0, \quad t_2 = \frac{i}{NT} T, \quad \begin{matrix} i = 1, 2, \ldots, NT-1 \\ i \neq NTAU \end{matrix} \tag{68}$$

so that the required numerical integrations can be performed. The numerical integration uses the trapezoidal rule. A more detailed description of these computations is given in Figure 9. The values of $V_o(\tau)$ and $V_o(T)$ are used in Block 5 in calculating $V(T,\tau)$.

The final set of computations is given in Block 6. First the steady-state covariance of the states is computed from the indicated equation, which is equation (51). The technique used is called "fast partial sums"; it is described in reference 4 and is incorporated in the DIGIKON subroutine MODCAL. Roughly, the technique is a modified form of successive substitution of the present value of $P_x(k)$ into (49) to compute $P_x(k+1)$; the modification enables n iterations to accomplish 2^n successive substitutions. Once P_{xss} is obtained, P_{e1ss} and P_{e2ss} are computed and the program is finished.

3. INSTRUCTIONS FOR USING THE PROGRAM

The user of the program must first check the dimensions of his problem to determine whether the array dimensions used are adequate for his problem. The program uses the technique of setting the maximum dimensions of all major arrays within DIMENSION statements in the main program. This permits adjusting array sizes by changes in the main program only.

Define

NPM = maximum number of plant states
NUPM = maximum number of plant control inputs
NWPM = maximum number of plant disturbance inputs
NOPM = maximum number of plant outputs
NCM = maximum number of controller states (69)

These variables are given numerical values by the first five FORTRAN arithmetic statements in the main program. Currently they are assigned the values 4, 2, 2, 3, and 2, respectively.

The two DIMENSION statements in the main program must use these same numerical values in the arrays. Table 1 shows what values to assign to the given arrays. In this table, the numerical values of NHM, NFM, and NRPM are calculated from

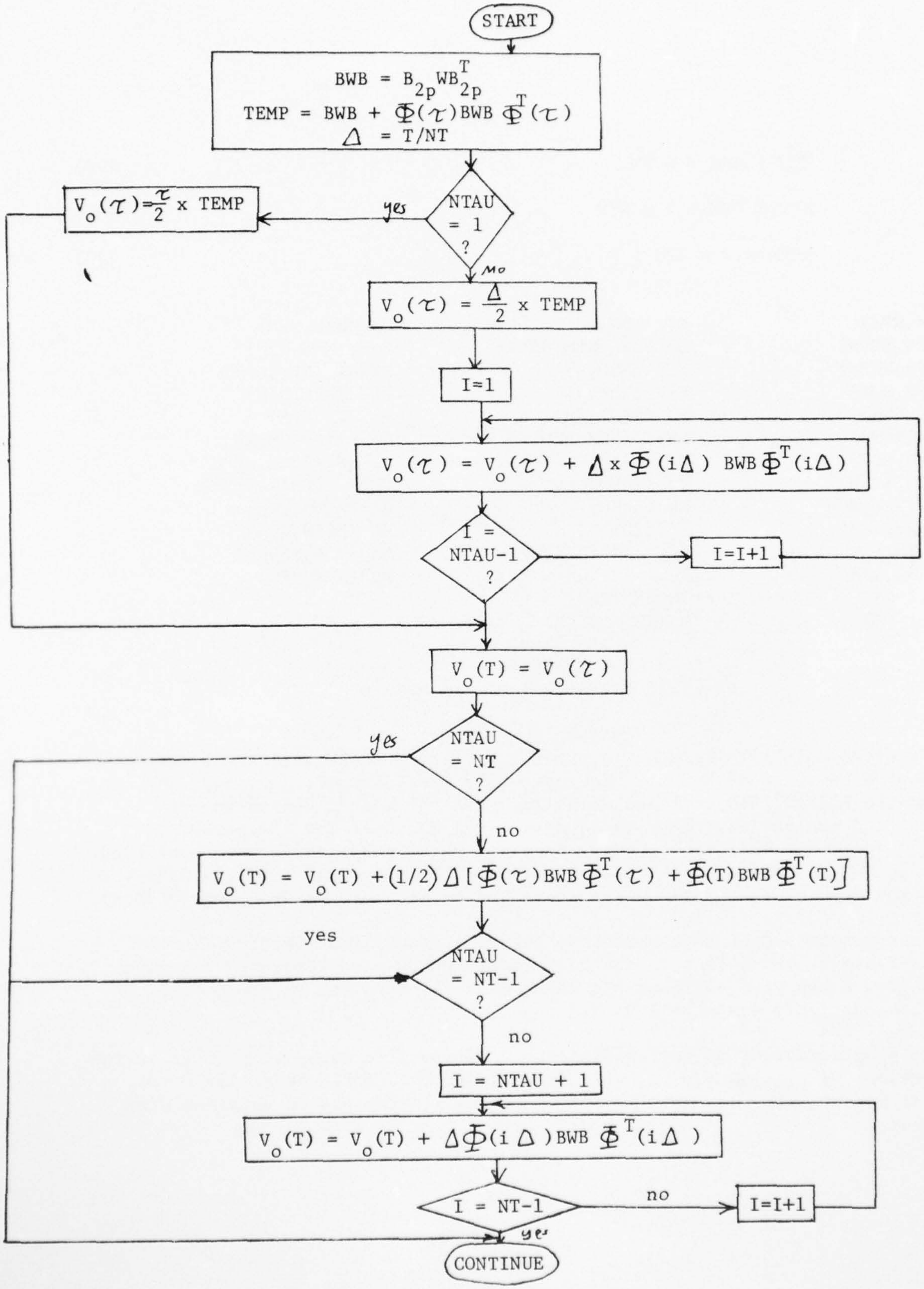

START
$BWB = B_{2p} W B_{2p}^T$
$TEMP = BWB + \Phi(\tau) BWB \Phi^T(\tau)$
$\Delta = T/NT$
NTAU = 1 ?
yes
$V_o(\tau) = \frac{\tau}{2} \times TEMP$
no
$V_o(\tau) = \frac{\Delta}{2} \times TEMP$
I=1
$V_o(\tau) = V_o(\tau) + \Delta \times \Phi(i\Delta)\ BWB\ \Phi^T(i\Delta)$
I = NTAU-1 ?
I=I+1
$V_o(T) = V_o(\tau)$
NTAU = NT ?
yes
no
$V_o(T) = V_o(T) + (1/2)\Delta[\Phi(\tau) BWB \Phi^T(\tau) + \Phi(T) BWB \Phi^T(T)]$
yes
NTAU = NT-1 ?
no
I = NTAU + 1
$V_o(T) = V_o(T) + \Delta\Phi(i\Delta) BWB\ \Phi^T(i\Delta)$
I = NT-1
no
I=I+1
yes
CONTINUE

Figure 9.
Flowchart For The Computation of $V_o(\tau)$ and $V_o(T)$

$$\mathrm{NHM} = \mathrm{NPM} + \mathrm{NUPM} \tag{70}$$

$$\mathrm{NFM} = \mathrm{NPM} + 2 \times \mathrm{NCM} \tag{71}$$

$$\mathrm{NRRM} = 2 \times \mathrm{NPM} + 4 \tag{72}$$

AP (NPM,NPM)
B1P (NPM,NUPM)
B2P (NPM,NWPM)
CP (NOPM,NPM)
FC (NCM,NCM)
GC (NCM,NOPM)
HC (NUPM,NCM)
EC (NUPM,NOPM)
PHITAU (NHM,NHM)
PHIT (NHM,NHM)
PSITAU (NHM,NHM)
PSIT (NHM,NHM)
F (NFM,NFM)
AH (NHM,NHM)

DD (NHM
DELPHI (NHM,NHM)
EIP (NHM,4)
EIV (NHM,4)
FH (NHM,NHM)
FHST (NHM,NHM)
KWA (NHM)
RF (NRRM)
RR (NRRM)
ID (20)
ECCP (NUPM,NPM)
PT (NHM,NHM)
INDEX (NHM)
W (NWPM,NWPM)

PS (NHM,NHM)
BWB (NPM,NPM)
VZTAU (NPM,NPM)
VZT (NPM,NPM)
V (NFM,NFM)
PXSS (NFM,NFM)
PE1SS (NUPM,NUPM)
PE2SS (NUPM,NUPM)
H1 (NUPM,NFM)
H2 (NUPM,NFM)
AM (NFM,NFM)
PM (NFM,NFM)
D (NFM,NFM)

Table 1
Required Dimensions Of All Arrays

The first data card is used to provide a message which will be printed at the top of a new page of output. The next card specifies NP, NUP, NWP, NOP, and NC using the FORMAT (5I3). These quantities are the actual dimensions of the plant and controller (refer to equation (69) but omit the adjective maximum). Next the matrices A_p, B_{1p}, B_{2p}, C_p, F_c, G_c, H_c, and E_c are specified in succession, one row and one card at a time, using the FORMAT (F10.4,2I3). Finally the rows of the matrix W are specified on separate cards using (8F10.4).

The complete set of data cards for the first example of Section IV are given in Table 2. Each line in the table corresponds to a separate data card. The quantity k has been assigned the value 0.5; T, NTAU, and NT are 1.0, 5, and 10, respectively and W is 1.0.

The program output is self-explanatory. A specific example is given in the next Section. A program listing will be included in a forthcoming Air Force Technical Report. In the interim, copies of the listing may be obtained from the author.

```
col. 12...

EXAMPLE 1

 1  1  1  1  1
0.0
1.0
1.0
1.0
0.0
0.0
0.0
-0.5
1.0         5 10
1.0
```

Table 2
Data For The First Example Of Section IV

SECTION VI

APPLICATION TO THE A7D PITCH AXIS FLIGHT CONTROL SYSTEM

1. CLOSED-LOOP MODEL

The major components of the closed-loop model used here for the A7D tactical fighter are the pilot-input, the aircraft, and the actuator dynamics, and the dual-redundant asynchronous controllers. The pilot-input dynamics, the aircraft dynamics, and the actuator dynamics are combined to form the plant model.

The pilot longitudinal-stick input measured in pounds of pressure is modeled as the response of a linear filter to a zero-mean white noise input. The filter is selected so that the frequency spectrum of its output closely approximates that of a typical pilot input. (See Reference 3, pages 55ff.) The selected filter is $\frac{\sqrt{2}\,\omega_0\sigma_w}{s+\omega_0}$ where σ_w is the root-mean-square command magnitude and ω_0 is the bandwidth. This filter can be expressed in state variable form as follows:

$$\dot{x}_3 = -\omega_0 x_3 + \sqrt{2}\,\omega_0\sigma_w w_p \tag{73}$$

where w_p is the filter input, and x_3 is the filter state and also the filter output.

The filter output is the pilot input to the redundant controllers. Using this formulation, x_3, the pilot input, becomes a state variable of the plant.

The aircraft dynamics are described by the equations

$$\dot{\alpha} = Z_w \alpha + \left(1 + \frac{Z_q}{u_o}\right)\dot{\theta} + \frac{Z_{\delta e}}{u_o}\delta_e \tag{74}$$

$$\ddot{\theta} = M_w u_o \alpha + M_{\dot{w}} u_o \dot{\alpha} + M_q \dot{\theta} + M_{\delta e}\delta_e \tag{75}$$

where α is the aircraft angle of attack expressed in radians, $\dot{\theta}$ is the pitch rate expressed in radians/second and δ_e is the elevator surface deflection from the nominal trim value measured in radians. The other quantities in (74) and (75) are stability derivatives whose values are given in Table 3 for the selected straight-and-level flight condition with the nominal forward velocity u_o = 334.9 feet/second at sea level.

Values of Model Coefficients

$Z_w = -.7572$

$\frac{Z_{\delta e}}{u_o} = -.1090$

$1 + \frac{Z_q}{u_o} = 1$

$M_w u_o = -3.764$

$M_{\dot{w}} u_o = -.08292$

$M_q = -.4682$

$M_{\delta e} = -6.083$

Table 3

A7D Stability Derivatives for u_o = 334.9 ft./sec
Nominal Forward Velocity at Sea Level

Let the actuator dynamics be described by a first-order lag transfer function $\frac{20}{s+20}$. These dynamics can be expressed in state variable form as follows:

$$\dot{\delta}_e = -20\,\delta_e + 20\,u_p \tag{76}$$

where u_p is the input to the actuator expressed in radians and δ_e is the actuator output which is also the aircraft elevator surface deflection. Using this formulation, δ_e becomes a state variable of the plant.

Define x_p by

$$x_p = \begin{bmatrix} \alpha \\ \dot{\theta} \\ x_3 \\ \delta_e \end{bmatrix} \tag{77}$$

Putting the pilot input, aircraft dynamics, and actuator dynamics equations in state variable form and using the values from Table 3, one obtains

$$\dot{x}_p = A_p x_p + B_{1p} u_p + B_{2p} w_p \tag{78}$$

where

$$A_p = \begin{bmatrix} -.7572 & 1.0 & 0 & -.109 \\ -3.7012 & -.5511 & 0 & -6.0734 \\ 0 & 0 & -\omega_o & 0 \\ 0 & 0 & 0 & -20.0 \end{bmatrix} \tag{79}$$

$$B_{1p} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 20 \end{bmatrix} \qquad B_{2p} = \begin{bmatrix} 0 \\ 0 \\ \sqrt{2}\,\omega_o \sigma_n \\ 0 \end{bmatrix} \tag{80}$$

The description of the digital controllers is obtained by starting with the Laplace transform transfer function of an analog controller and performing the Tustin transformation (also called the bilinear transformation) to obtain the Z transfer function and the discrete-time state equations.

The continuous A7D pitch-axis control law for each controller is illustrated in Figure 10 where $[N_z - 1]$ is the output in g's of the normal accelerometer placed seven feet forward from the center of gravity, $\dot{\theta}$ is the output of the pitch-rate gyro in radians/second, x_3 is the pilot longitudinal-stick input in pounds with positive force measured in the aft direction.

Sensor dynamics have not been included in the plant dynamics to simplify the analysis.

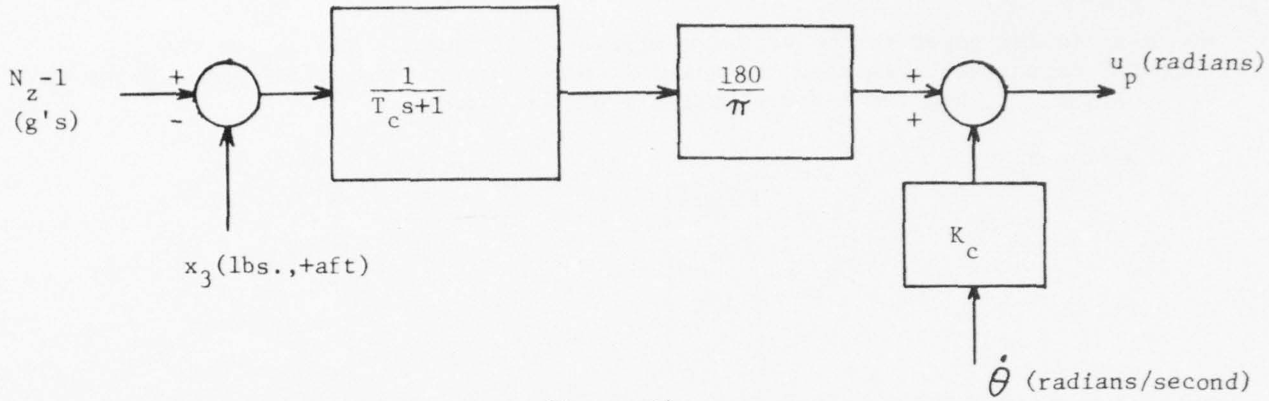


Figure 10.
Block Diagram of the Continuous A7D Pitch-Axis Control Law

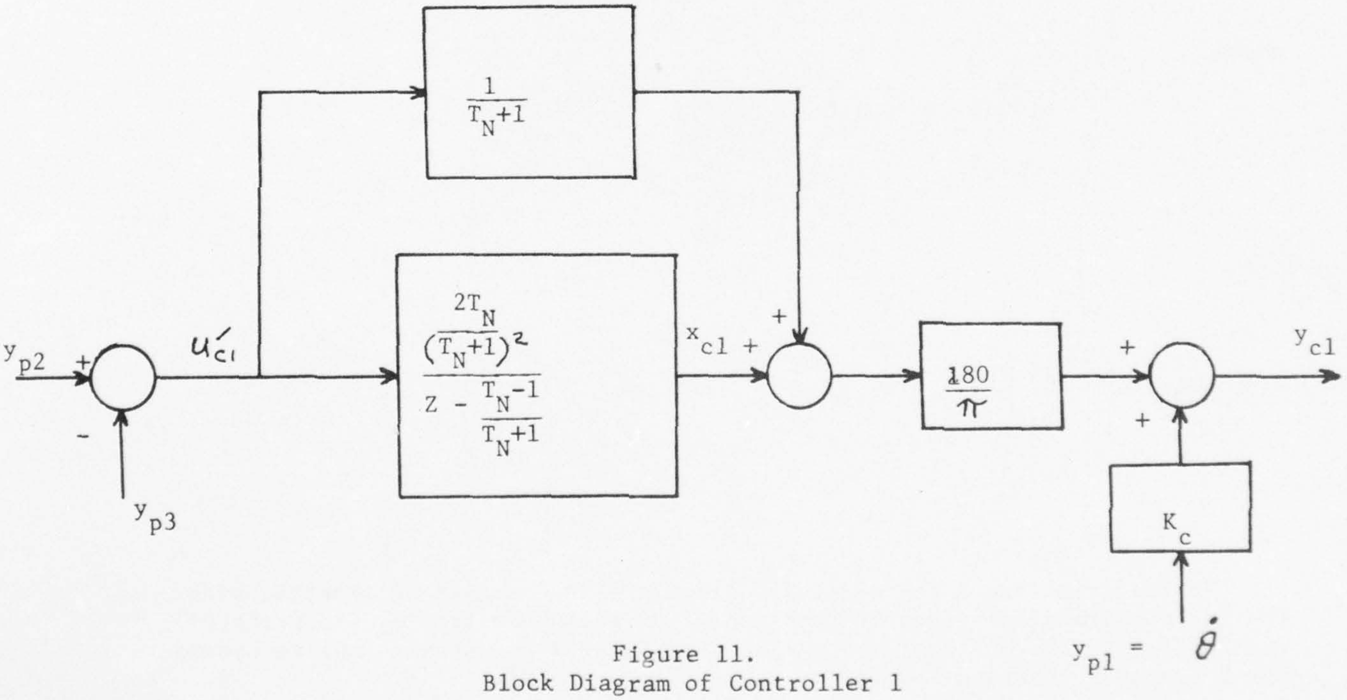
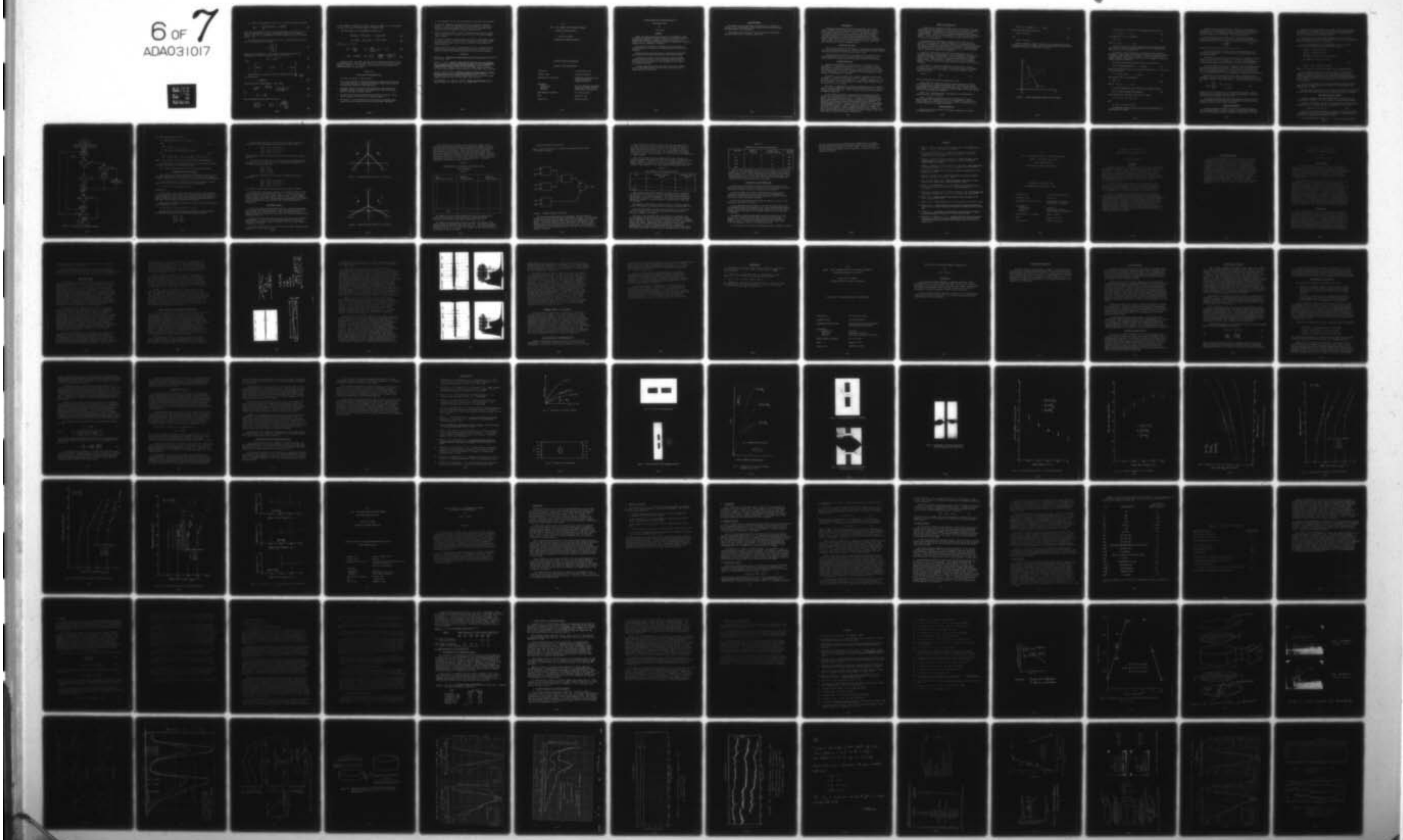
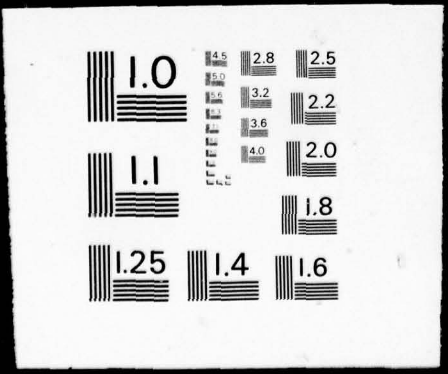


Figure 11.
Block Diagram of Controller 1

AD-A031 017 AUBURN UNIV ALA SCHOOL OF ENGINEERING F/G 5/2
1975 USAF/ASEE SUMMER FACULTY RESEARCH PROGRAM.(U)
SEP 75 J F O'BRIEN F44620-75-C-0031
UNCLASSIFIED AFOSR-TR-76-1113 NL

6 OF 7
ADA031017




1.0
4.5
5.0
5.6
6.3
7.1
8.0
9.0
10
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

For flight near straight and level, $[N_z - 1]$ can be expressed as follows:

$$N_z - 1 = \frac{1}{g}\left[u_o(\dot{\theta} - \dot{\alpha}) + \ell_x \ddot{\theta}\right] \tag{81}$$

where g is the acceleration due to gravity expressed in ft./sec^2, and 1_x is 7 feet, the placement of the normal accelerometer forward of the center of gravity. Also, for the standard A7D pitch-axis control law $T_c = 0.55$ seconds and $K_c = 0.25$.

The output of the plant is taken as the set of quantities which are fed back to the controllers. These are

$$y_p = \begin{bmatrix} \dot{\theta} \\ N_z - 1 \\ x_3 \end{bmatrix} \tag{82}$$

where the second component is the normal acceleration of equation (81). Expressing y_p in terms of the plant state x_p, one obtains

$$y_p(t) = C_p x_p(t) \tag{83}$$

where

$$C_p = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 7.07074 & -0.11981 & 0 & -0.18676 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{84}$$

Consider now the continuous controller transfer function $\frac{180}{\pi} \cdot \frac{1}{T_c s+1}$. The substitution

$$s = \frac{2}{T}\frac{z - 1}{z + 1} \tag{85}$$

performs the Tustin transformation. This yields

$$\frac{1}{T_c s + 1} \xrightarrow{Tustin} \frac{1}{T_c \frac{2}{T} \frac{z-1}{z+1} + 1} \tag{86}$$

Let

$$T_N = 2\frac{T_c}{T} \tag{87}$$

Then the transfer function can be written as

$$\frac{X_{c1}(z)}{U'_{c1}(z)} = \frac{1}{T_N + 1} + \frac{\frac{2T_N}{(T_N+1)^2}}{z - \frac{T_N - 1}{T_N + 1}} \tag{88}$$

A block diagram for digital controller 1 appears in Figure 11. In this Figure y_{p1}, y_{p2}, and y_{p3} are the plant outputs of equation (82).

The state equations corresponding to Figure 11 are

$$x_{c1}(t_{k+1}) = F_c x_{c1}(t_k) + G_c u_{c1}(t_k) \tag{89}$$

$$y_{c1}(t_k) = H_c x_{c1}(t_k) + E_c u_{c1}(t_k) \tag{90}$$

where $t_{k+1} - t_k = T$ and

$$F_c = \frac{T_N - 1}{T_N + 1} \qquad G_c = \frac{2T_N}{(T_N+1)^2} \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} \tag{91}$$

$$H_c = 57.296 \qquad E_c = \begin{bmatrix} K_c & \frac{57.296}{T_N + 1} & \frac{-57.296}{T_N + 1} \end{bmatrix} \tag{92}$$

Equations (78), (83), (89), and (90) give the required description of the closed-loop system. They were used to provide the input data for the software package of Section VI. The results of this analysis will be documented in a forthcoming AFFDL Technical Report.

SECTION VII

CONCLUSIONS AND RECOMMENDATIONS

The major conclusions of this work are

1. It is indeed possible to develop an analytical formulation to model the closed-loop operation of asynchronous redundant digital flight controls incorporating reasonable voting algorithms.

2. Attendant covariance analyses can be developed to characterize in a meaningful statistical manner the inherent errors that exist between pairs of redundant controller outputs.

3. The models and covariance analyses can be applied to realistic flight control problems with the aid of the digital computer.

4. The results of limited application of the analytical treatment agree qualitatively with engineering intuition and hardware simulations.

It is recommended that the work be continued to accomplish the following:

1. Increase the complexity and generality of the models and covariance analyses to include such features as multirate sampling, computational delays, processor word-length effects, sensor noise, additional voter algorithms, and new measures of inherent channel errors.

2. Develop packaged software to permit the application of the new formulations to the analysis of realistic and practical asynchronous redundant flight controls.

3. Use the new software to perform parametric studies of applicable systems of current interest to the Air Force. Parameters include control system gains and time constants, sample periods, processor word lengths and flight conditions.

4. Correlate the results with the AFFDL Digital Avionics Information System (DAIS) laboratory hardware simulations and develop a laboratory test program to evaluate the hazards associated with asynchronous operation.

REFERENCES

1. Meditch, J.S., Stochastic Optimal Linear Estimation and Control, McGraw-Hill Book Company, 1969.

2. Konar, A.F., et al. Digital Flight Control Systems For Tactical Fighters, Vol. I: Digital Flight Control Systems Analysis, Vol. II: Documentation of the Digital Control Analysis Software (DIGIKON), Technical Report AFFDL-TR-73-119, Volumes I and II, Air Force Flight Dynamics Laboratory, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, June 1974.

3. Stein, G. and A.H. Henke, A Design Procedure And Handling-Quality Criteria For Lateral-Directional Flight Control Systems, Technical Report AFFDL-TR-70-152, Air Force Flight Dynamics Laboratory, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, May 1971.

4. Van Dierendock, A.J., and G.L. Hartmann, Quadratic Methodology, Vol. I, Honeywell Report No. F0161-FR, Vol. 1, October, 1973.

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT-PATTERSON AFB,OHIO

&

EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

PATTERN RECOGNITION TECHNIQUES

APPLIED TO FLAT-BOTTOM HOLES

Prepared by:	J. Kent Bryan, Ph.D.
Academic Rank:	Assistant Professor
Department and University:	Department of Electrical and Computer Engineering Clemson University
Assignment: (Laboratory) (Division) (Branch)	 Air Force Materials Laboratory Metals and Ceramics Division Nondestructive Evaluation Br.
USAF Research Colleague:	M.J. Buckley, Ph.D.
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

PATTERN RECOGNITION TECHNIQUES APPLIED TO FLAT-BOTTOM HOLES

By

J. Kent Bryan

ABSTRACT

Signal processing and pattern recognition techniques have been increasingly recognized as important factors in the design of modern computerized information systems. They have been used in such varied fields as biology, psychology, medicine, engineering, statistics, computer science, chemistry, and physics.

This study was conducted to investigate the applicability of pattern recognition and signal processing methods to nondestructive evaluation.

Excellent results were obtained in discriminating flat-bottom hole sizes based on digitized ultrasonic pulse echo waveforms.

The linear discriminant function and the adaptive learning network procedures are very easy to implement and both correctly identified 48 of 49 samples for a 98% recognition rate based on 15 waveform parameters.

A further investigation indicated that all of the 49 samples may be dichotomized correctly if the proper coefficients for a linear function are specified.

ACKNOWLEDGEMENTS

The author sincerely expresses his gratitude to Dr. Michael J. Buckley and all of the members of the Nondestructive Evaluation Branch for their support and encouragement.

The assistance and patience of Mrs. Kayann Cornish during the manuscript preparation is sincerely appreciated.

INTRODUCTION

Ultrasonic nondestructive testing has been widely used for many years. Classically it has been used to determine physical or crack like flaws in metallic structures. Because of the inherent capability of ultrasonic instruments to measure both the amplitude and phase of reflected signals, it is being examined for applicability of signal processing techniques to extract extra information from the reflected signal. These signal processing techniques have been used by communication engineers for some time to extract information from a signal that would otherwise be buried in background noise.

OBJECTIVES AND SCOPE

The principal objective of this work is to investigate the applicability of pattern recognition procedures to ultrasonic pulse echo waveforms obtained from test blocks with different flat-bottom hole sizes.

Pattern recognition literature is reviewed and some techniques which can be implemented by the use of computer programs are explored. The results based on artificial data and ultrasonic waveforms are discussed.

PATTERN RECOGNITION

Automated interpretation of ultrasonic pulse echo waveforms can be formulated as a classification problem in which it is desired to decide into which of M categories denoted by $C_1, \ldots, C_m$, each test block belongs. The numerical results from preprocessing time waveforms for each test block can be represented by a point in n-dimensional space assuming that results consist of n measurements or features.

A pattern is a n-tuple where each component represents a measurement or feature. A pattern is denoted by the vector X where $X = (x_1, \ldots, x_n)$. Let R^n denote the set of all n-tuples $(x_1, \ldots, x_n)$ whose components $x_1, \ldots x_n$ are real numbers. R^n is called n-space, and each n-tuple in R^n is said to be a point or vector in n-space.

The topic of classification is usually included under the more general topic of pattern recognition. Several good survey papers (Ref. 1,2,3,4) and recently published books (Ref. 5,6,7) have been written describing pattern recognition techniques.

In most pattern recognition problems the only information available consists of a "training" set of N patterns whose true classifications are known. The N patterns denoted by $X_1, \ldots, X_N$ are called training samples. The training samples from each category are assumed to be independent and identically distributed according to some unknown density function. The training samples are used to construct decision rules which are implemented by discriminant functions. These functions are defined to be real-valued functions of the pattern X used in classifying pattern X as a member of one of the M categories. The discriminant functions yield a decision rule which specifies that pattern X is classified as being a member of that class which has the largest discriminant functionvalue.

SAMPLE AND FEATURE SIZE

In most pattern recognition problems little is known about the underlying probability distributions of the M categories or classes. Therefore, the discriminant functions must be determined on the basis of the N training samples. The classification results obtained from the training samples should be related to the performance of the decision rule on future samples. Quite often the error rate obtained is lower than the true error rate of the classifier.

One method quite often used to estimate the true error rate of a classifier is to divide the original N samples into a design set and a test set. The design set is then used as the training set and the resulting discriminant functions are then tested on the test set. One problem with this approach is that the value of N may be small and a better classifier could be designed by using all of the samples.

Based on a fixed sample size Kanal and Chandrasekaran (Ref. 8) recommend using the "leaving-one-out" method in designing a classification system and evaluating its performance. In the leaving-one-out method a classifier is designed based on N-1 samples and then tested on the one removed sample. This procedure is repeated for each of the N training samples. A problem associated with this approach is that it may be too time consuming.

Foley (Ref. 9) using both experimental and theoretical results, indicates that the ratio of the number of samples per class to the number of features should be at least three to obtain good estimates of the optimum error rate.

That is,

$$\frac{N_1}{n} \geq 3 \qquad (1)$$

where N_1 represents the number of samples per class.

Meisel (Ref. 10) points out that the n used in equation (1) should in some sense be the "intrinsic dimensionality." This means that the set of n features should contain only relevant information. Feature selection schemes (Ref. 4,7, 10) might be used in reducing n-space to one that contains only useful discriminatory information.

Kanal (Ref. 4) discusses other investigations into dimensionality, sample size, and error estimation.

When a classifier's performance cannot be generalized to future unknown samples the classifier is said to be overtrained or overfitted. This situation occurs most frequently when the number of samples per class is small compared to the number of features.

LINEAR SEPARABILITY

Two pattern classes are said to be linearly separable if a linear discriminant function

$$d(X) = w_0 + w_1x_1 + w_2x_2 + \ldots\ldots\ldots\ldots + w_nx_n \qquad (2)$$

exists which has the property that

$$d(X) > 0 \text{ for } X\varepsilon C_1$$

and (3)

$$d(X) < 0 \text{ for } X\varepsilon C_2$$

Figure 1 displays a linear classifier for two hypothetical pattern classes in 2-space. In general, the linear discriminant function defined by equation (2) defines a hyperplane in n-space.

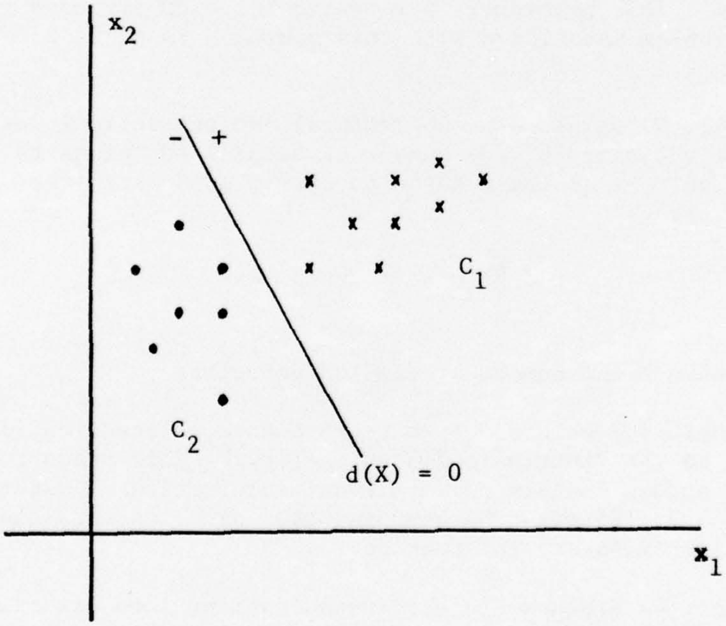


FIGURE 1. LINEAR DISCRIMINANT FUNCTION FOR TWO CLASSES

If the vector Y is defined as the augmented pattern vector X

$$Y = (1,x_1,x_2.........,x_n) \tag{4}$$

then equation (2) can be expressed as

$$d(Y) = W{\cdot}Y \tag{5}$$

where $W = (w_0,w_1, ...,w_n)$.

In order to specify a linear discriminant function which separates two classes in n-space the training samples are used to determine the values for the n+1 coefficients of W.

In practice, pattern classes are not usually linearly separable. Although, in order to gain the advantage of simplicity, one may be willing to sacrifice some performance with regard to correct classification of training samples. This linear concept can also be extended to cover more complex, nonlinear boundaries.

The generalized linear discriminant function has the form

$$d(X) = w_0+w_1f_1(X)+w_2f_2(X)+.........+w_Lf_L(X) \tag{6}$$

where the L functions $f_i(X)$ can be arbitrary functions of X. If the vector Y is now defined as

$$Y = (1,f_1(X), f_2(X), ...,f_L(X)) \tag{7}$$

then equation (6) can be expressed as

$$d(Y) = W{\cdot}Y \tag{8}$$

where $W = (w_0, w_1, ..., w_L)$

In the two-dimensional case a quadratic discriminant function

$$d(X) = w_0+w_1x_1+w_2x_2+w_3x_1^2+w_4x_1x_2+w_5x_2^2 \tag{9}$$

can be expressed in the linear form $d(Y) = W{\cdot}Y$ where

$$Y = (1, x_1, x_2, x_1^2, x_1x_2, x_2^2)$$

and

$$W = (w_0, w_1, w_2, w_3, w_4, w_5).$$

The nonlinear problem in 2-space has now been transformed into a linear problem in 5-space.

The generalized linear discriminant function can thus be seen to correspond to a linear problem in L-space, although the L functions $f_i(X)$ may be nonlinear. The discriminant function also retains its general nonlinear properties in the n-space of the original patterns.

As can be seen above six terms are involved in describing the quadratic discriminant function in 2-space. In general, the number of terms needed to describe a p-th degree polynomial in n-space is given by

$$N_T = \frac{(n+p)!}{p!n!} \qquad (10)$$

which grows quite rapidly as a function of n and p.

It is not necessary to use all of the N_T terms though since some of them may be unimportant in dichotomizing the training samples into two class groupings.

As mentioned in the last section, the ratio of the number of training samples N to the number of features n has an effect on the estimate of the true error rate of the classifier. In terms of linear dichotomies of the training samples the relationship of N to n is also important.

In the following discussion it is assumed that the N training samples are in "general position" in n-space. For $N>n+1$ a set of N points is said to be in general position if no subset of n+1 of the N points lies on an (n-1)-dimensional hyperplane. For $N\leq n+1$ a set of N points is said to be in general position if no (N-2)-dimensional hyperplane contains the set.

Since there are two classes, the total number of dichotomies (not necessarily linear) of the N points is 2^N. It can be shown (Ref. 11) that the number of distinct linear dichotomies of the N points in n-space is given by

$$L(N,n) = \begin{cases} 2 \sum\limits_{i=0}^{n} \binom{N-1}{i}, & \text{for } N>n+1 \\ 2^N, & \text{for } N\leq n+1 \end{cases} \qquad (11)$$

Hence, it can be concluded that if $N \leq n+1$ then there exists a linear discriminant function which effects the same dichotomization as specified by the two class assignment of the N points.

This situation demonstrates that overtraining may occur when $N\leq n+1$ since a separator does exist that will dichotomize the samples correctly, although it may not be the optimal classifier.

PERCEPTRON ALGORITHM

A class of machines developed as a model of machine learning and decision making has been called a perceptron and has played an important role in the development of pattern recognition theory (Ref. 6, 12, 13).

The basic perceptron algorithm is a simple scheme for the iterative determination of the weight vector W used to define the hyperplane $d(Y) = W\cdot Y = 0$, which is a linear discriminant function. An outline of the perceptron algorithm can be stated as follows.

Given two sets of training samples belonging to pattern classes C_1 and C_2, respectively, let the initial weight vector W_1 be chosen arbitrarily. Then, the (K+1)st approximation is given by:

1. If the Kth member of the training sequence Y_K is classified correctly leave the weight vector W_K unchanged. That is,

$$\begin{aligned} W_{K+1} &= W_K \quad \text{if } W_K\cdot Y_K > 0 \text{ and } Y_K \varepsilon C_1 \\ W_{K+1} &= W_K \quad \text{if } W_K\cdot Y_K < 0 \text{ and } Y_K \varepsilon C_2 \end{aligned} \tag{12}$$

2. Otherwise, the weight vector is changed by

$$W_{K+1} = W_K + cY_K \quad \text{if } W_K\cdot Y_K \leq 0 \text{ and } Y_K \varepsilon C_1$$

or (13)

$$W_{K+1} = W_K - cY_K \quad \text{if } W_K\cdot Y_K \geq 0 \text{ and } Y_K \varepsilon C_2$$

where c is a positive correction increment, possibly depending upon K.

The algorithm is said to have converged when all of the training samples are classified correctly. It can be shown that if the two classes are linearly separable then the perception algorithm converges in a finite number of iterations (Ref. 6, 7, 11). The correction increment c may be selected in several ways although in practice a value of c = 1 works quite well.

A flow chart describing the perceptron algorithm is given in Figure 2. When the value of MCNT equals the total number of training samples the algorithm has converged.

This perceptron algorithm can also be used to determine the coefficients of non-linear discriminant functions as was mentioned in the last section.

MULTICATEGORY CLASSIFICATION

The basic perceptron algorithm described above is based on the two class problem, although it is easily extended to the multiclass situation.

For the M class problem the idea is to determine M linear discriminant functions $d_1(Y),......,d_M(Y)$, with the property that if $Y \varepsilon C_i$, then

$$d_i(Y) > d_j(Y) \quad \text{for all } j \neq i \tag{14}$$

When all of the training samples are classified correctly by equation (14) the M classes are said to be linearly separable. A generalization of the perceptron algorithm is described as follows:

Let the initial weight vectors $W_1,........,W_M$ be chosen arbitrarily.

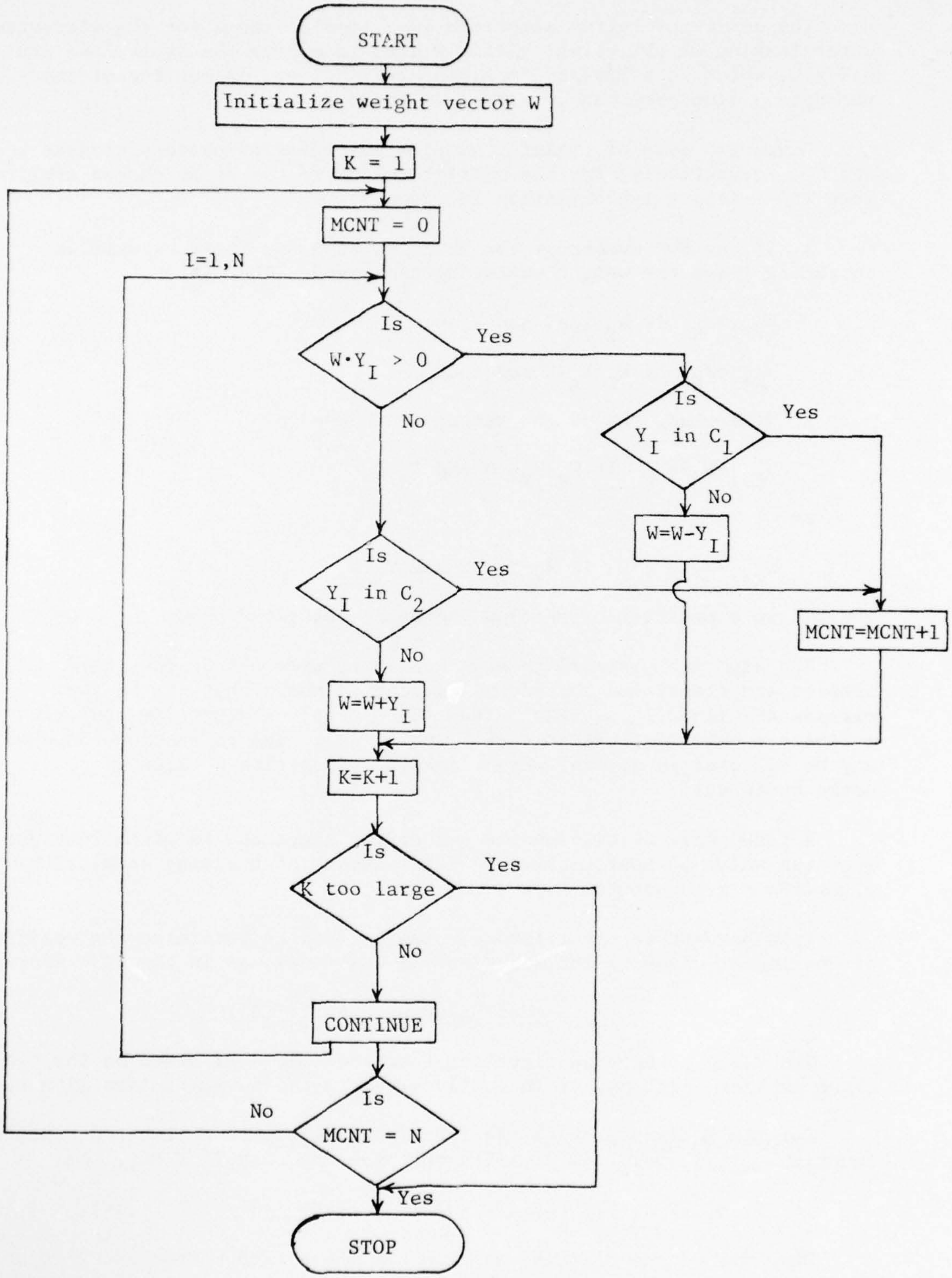


FIGURE 2. FLOW CHART OF PERCEPTRON ALGORITHM

The (K+1)st approximation is given

1. If $W_i^K \cdot Y_K > W_j^K \cdot Y_K$ for all $j \neq i$ and $Y_K \varepsilon C_i$,

then (15)

$$W_t^{K+1} = W_t^K \text{ for } t = 1, \ldots\ldots, M$$

2. If $W_t^K \cdot Y_K > W_i^K \cdot Y_K$ and $W_i^K \cdot Y_K > W_j^K \cdot Y_K$ for $Y_K \varepsilon C_i$,

then (16)

$$W_t^{K+1} = W_t^K - cY_K,\ W_i^{K+1} = W_i^K + cY_K, \text{ and } W_j^{K+1} = W_j^K \text{ for } j \neq t,i$$

where c is a positive correction increment, possibly depending upon K.

If the M classes are linearly separable then it can be shown that this algorithm converges in a finite number of steps.

The following example illustrates the results of this procedure.

AN EXAMPLE WITH ARTIFICIAL DATA

This example is not being considered to demonstrate the utility of the perceptron algorithm for a multiclass problem, but instead as an example to verify the theoretical results stated in the section on Linear Separability.

Tou and Gonzalez (Ref. 7) presented a three class example in 2-space where each class C_1, C_2, and C_3 contains one training sample,

$$X_1 = (0,0),\ X_2 = (1,1), \text{ and } X_3 = (-1,1),$$

respectively.

In this example n is equal to three and N is equal to two for each of the three pairs of classes. Based on the results stated in the section on Linear Separability, linear discriminant functions exist which separate each pair of classes since $N \leq n+1$ and no (N-2)-dimensional hyperplane (point in this case) contains each of the three pairs of points.

Before the multicategory algorithm can be applied the training samples must be augmented to yield

$$Y_1 = (1,0,0),\ Y_2 = (1,1,1), \text{ and } Y_3 = (1,-1,1).$$

Starting with all three initial weight vectors set equal to (0,0,0) the algorithm converges to the following discriminant functions.

$$d_1(X) = -2x_2$$
$$d_2(X) = 2x_1-2$$
$$d_3(X) = -2x_1-2$$

The decision boundaries for these results are found by setting the difference of these discriminant functions equal to zero. That is,

$$d_1(X) - d_2(X) = 2-2x_1-2x_2 = 0$$
$$d_1(X) - d_3(X) = 2+2x_1-2x_2 = 0$$
$$d_2(X) - d_3(X) = 4x_1 = 0$$

These boundaries define three decision regions for each class as shown in Figure 3.

If each initial weight vector is set equal to the training sample in its class (i.e., $W_1 = Y_1, W_2 = Y_2, W_3 = Y_3$) the algorithm converges to the following discriminant functions.

$$d_1'(X) = 2-2x_2$$
$$d_2'(X) = 2x_1+2x_2$$
$$d_3'(X) = -2x_1+2x_2$$

The decision boundaries based on these discriminant function are as follows:

$$d_1'(X) - d_2'(X) = 2-2x_1-4x_2 = 0$$
$$d_1'(X) - d_3'(X) = 2+2x_1-4x_2 = 0$$
$$d_2'(X) - d_3'(X) = 4x_1 = 0$$

The decision regions formed by these boundaries are shown in Figure 4.

The decision regions shown in Figures 3 and 4 indicate that the final results depend upon the initial weight values. At this point there is no way to tell why one result is favorable over the other. As more samples become available these weights may be adjusted to obtain better estimates of their optimal values. This suggests that any generalization at this stage as to the probability of error for this three class problem would be inappropriate.

ULTRASONICS EXAMPLE

The Air Force Materials Laboratory initiated a program to determine if an advanced signal processing system could classify the ultrasonic pulse echo waveforms from flat-bottom holes. This study examined forty nine samples obtained from aluminum area-amplitude test blocks and three different transducers (Ref. 14).

Sixteen test blocks were fabricated from two different sets of 7075-T6 aluminum alloy. Each of the two sets contained eight test blocks which had flat-bottom hole sizes ranging in diameter from 1/64 to 8/64 inches in increments of 1/64-inch.

The three transducers used in this study were all 5MHz transducers with diameters of 0.5, 0.75, and 1.0 inch.

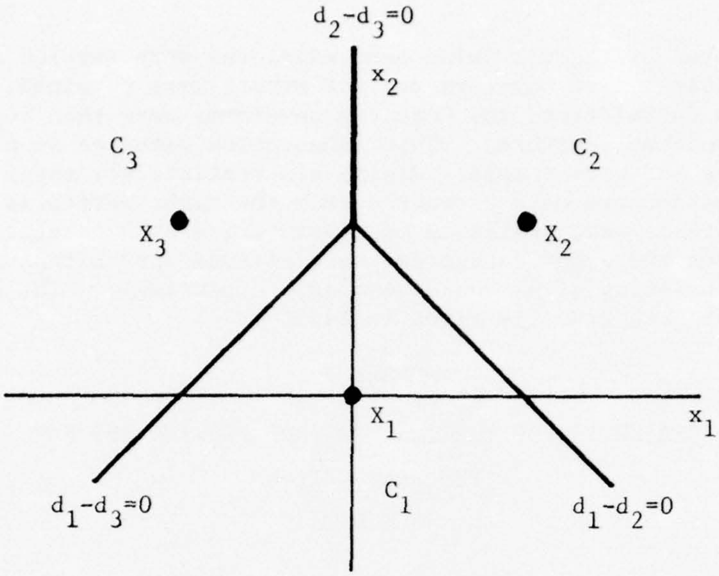


FIGURE 3. DECISION REGIONS BASED ON d_1, d_2, AND d_3

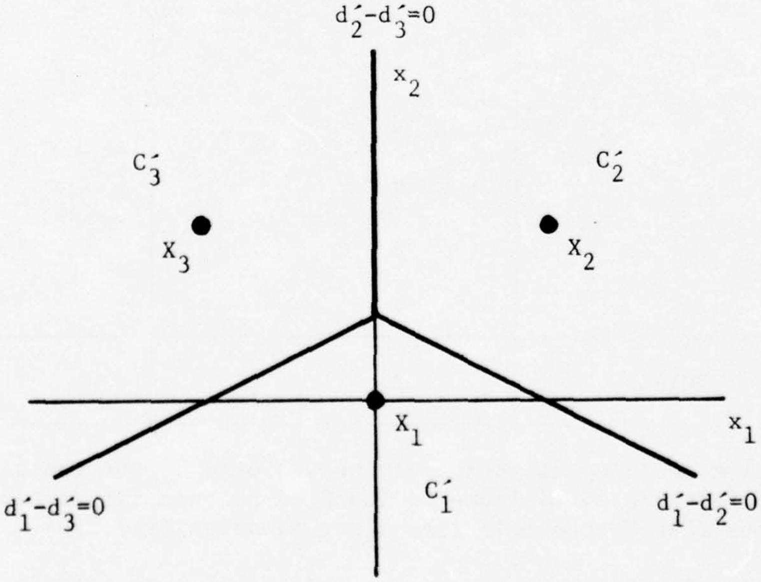


FIGURE 4. DECISION REGIONS BASED ON d_1', d_2', AND d_3'

The received ultrasonic pulse echo waveforms were sampled and the resulting digitized time waveform and its square were obtained. The Power Spectrum, Auto Correlation, and Cepstrum waveforms were then computed for each of the two time waveforms. This computation resulted in a total of eight waveforms for each sample. Ninety six statistical, shape, area and extremal parameters were extracted from the eight waveforms. Therefore, ninety six features were available to characterize each sample. The forty nine samples for the eight categories were divided into a training set and a testing set consisting of 41 and 8 samples, respectively. The class distribution of each class is given in Table I.

TABLE I

DISTRIBUTION OF THE TRAINING AND TESTING SET FOR

THE FLAT-BOTTOM

HOLE DATA

CLASS (Hole Sizes in 64th's)	NUMBER OF Training Samples	NUMBER OF Testing Samples
1	5	1
2	5	1
3	3	1
4	5	1
5	5	1
6	5	1
7	6	1
8	7	1
TOTAL	41	8

The adaptive learning network approach was used by the contractor to reduce the 96 features to 15 features found to contain information relative to the flat-bottom hole size discrimination (Ref. 14).

The adaptive learning network approach is similar in principal to the layered networks of threshold logic units (Ref. 11). The resulting learning network has the layered form illustrated in Figure 5. Each block implements a quadratic function of its inputs. For example, the block with inputs x_1 and x_{10} in Figure 5 has as its output

$$a_0+a_1x_1+a_2x_{10}+a_3x_1^2+a_4x_1x_{10}+a_5x_{10}^2$$

Thus, it can be seen that the resulting discriminant function d(X) is a 4th degree polynomial.

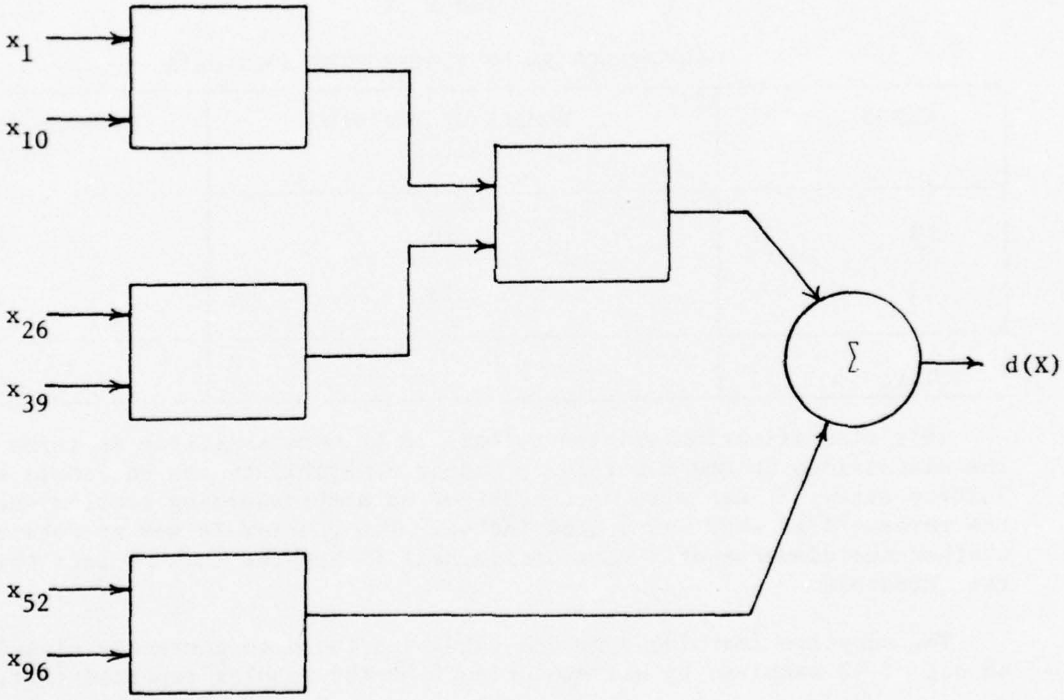


FIGURE 5. LEARNING NETWORK ILLUSTRATION

The adaptive learning network also implements a feature selection procedure to determine the important discriminatory features. This procedure involves dividing the training samples into a fitting subset and a selection subset. The fitting subset is used to determine the coefficients of the quadratic output for each block in the first layer. The selection subset is then used to reject the blocks with poor performance. This procedure is repeated for each succeeding layer until the error rate on the selection subset is minimized.

Based on the results stated in the section on Linear Separability, the eight classes are linearly separable. This was verified also by the use of the perceptron algorithm. The ratio of the number of training samples per class, which varies from 3 to 7, to the number of features (15) is less than 0.5 for each class. This ratio should be much higher as indicated earlier if the performance of a classifier is to be generalized to unknown samples.

The overall problem can be reformulated into a more satisfactory problem in terms of the relationship between the number of samples and the number of features. Classes 1,2,3, and 4 can be grouped together as a single category and classes 5,6,7 and 8 can be grouped together as a second category. This reduces to a two class problem as indicated in Table II.

TABLE II

TWO CATEGORIES OF FLAT-BOTTOM HOLE DATA

CLASS	NUMBER OF TRAINING SAMPLES	NUMBER OF TESTING SAMPLES
1	18	4
2	23	4
TOTAL	41	8

This classification problem appears to be more realistic in terms of the discussions in the sections on Linear Separability and on Sample and Feature size. It can also be thought of as a thresholding problem where the threshold is equal to 4.5/64 inches. The problem is now to determine whether the diameter of a flat-bottom hole is greater than or less than the threshold.

The adaptive learning approach (ALN) was found to correctly classify 48 out of 49 samples, by extrapolating from the results reported (Ref. 14).

The linear discriminant function approach (LDFI)correctly classified 48 out of 49 samples and a 3rd degree polynomial (POLY) was found to correctly classify all 49 samples.

All three of the procedures ALN, LDFI and POLY were developed using training samples and were tested on testing samples. These three results can be compared in terms of the number of coefficients required to implement each of the schemes. Based on the results shown in Table III the ALN and LDFI methods both require considerably fewer coefficients than the POLY procedure does. In terms of storage requirements on a computer or implementation in terms of hardware the ALN and LDFI methods would require less storage and perhaps would be faster to implement than the POLY method.

TABLE III

COMPARISON OF CLASSIFICATION PROCEDURES

PROCEDURE	DEGREE OF POLYNOMIAL	NUMBER OF CO-EFFICIENTS REQUIRED	PER CENT CORRECT
ALN	8	78	98%
LDFI	1	16	98%
POLY	3	816	100%
LDF2	1	16	100%

Another linear discriminant function procedure (LDF2) is also indicated in Table III. This procedure requires only 16 coefficients and achieves a 100% recognition rate. These coefficients were obtained by the perceptron algorithm using all of the forty nine samples in the training phase. These results indicate that all of the forty nine samples may be dichotomized correctly if the proper coefficients are specified.

CONCLUSIONS AND RECOMMENDATIONS

The object of this work has not been to advocate one procedure over another but instead to explore the possibility of using pattern recognition and signal processing techniques in nondestructive evaluation.

It has been shown with a small training sample size that pattern recognition techniques can be used effectively in discriminating flat-bottom hole sizes using signal processing techniques to generate digital data from ultrasonic pulse echo waveforms.

Pitfalls were indicated when the number of samples per class is small compared to the number of features.

A very simple approach was chosen to indicate that good results could be obtained with very little sophistication. Most researchers would agree that the linear separable technique is most likely not the optimal solution. A method such as this should not be ignored though since it is very easy to implement.

The adaptive learning network approach performs extremely well both as a pattern recognition procedure and a feature selection scheme. It appears to be a very powerful technique which should be investigated further in terms of its ability to generalize from small sample sizes to its discriminatory ability in the large sample case.

In the future, the use of Rome Air Development Center's (RADC) interactive pattern recognition facility OLPARS should be explored since it allows the researcher who understands the physical problems insert his knowledge into the solution by interaction at a CRT display. This facility has excellent waveform processing, feature extraction, and pattern recognition techniques available.

REFERENCES

1. Nagy, G., "State of the Art in Pattern Recognition," Proceedings of the IEEE, Vol. 56, No. 5, May 1968, pp. 836-862.

2. Ho, Y.C. and Agrawala, A.K., "On Pattern Classification Algorithms - Introduction and Survey," Proceedings of the IEEE, Vol. 56, No. 12, December 1968, pp. 2101-2114.

3. Wee, W.G., "A Survey of Pattern Recognition," IEEE Proceedings of the Seventh Symposium on Adaptive Processes, Los Angeles, California, December 1968, pp. 2e1-2e13.

4. Kanal, L., "Patterns in Pattern Recognition: 1968-1974," IEEE Transactions on Information Theory, Vol. IT-20, No. 6, November 1974, pp. 697-722.

5. Fukunaga, K., Introduction to Statistical Pattern Recognition, New York: Academic Press, 1972.

6. Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis, New York: John Wiley & Sons, Inc., 1973.

7. Tou, J.T. and Gonzalez, R.C., Pattern Recognition Principles, Reading, Massachusetts: Addison-Wesley Publishing Company, 1974.

8. Kanal, L. and Chandrasekaran, B., "On Dimensionality and Sample Size in Statistical Pattern Classification," Pattern Recognition, Vol 3, 1971, pp. 225-234.

9. Foley, D.H., "Considerations of Sample and Feature Size," IEEE Transactions on Information Theory, Vol. IT-18, No. 5, September 1972, pp. 618-626.

10. Meisel, W.S., Computer-Oriented Approaches to Pattern Recogntion, New York: Academic Press, 1972.

11. Nilsson, N.J., Learning Machines, New York: McGraw-Hill Book Company, 1965.

12. Rosenblatt, F., The Perceptron-A Perceiving and Recognizing Automation, Report 85-460-1, Cornell Aeronautical Laboratory, Ithaca, New York, January 1957.

13. Rosenblatt, F., Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms, Washington, D.C.: Spartan Books, 1962.

14. Mucciardi, A., Shankar, R., et al., Adaptive Nonlinear Signal Processing for Characterization of Ultrasonic NDE Waveforms, Air Force Materials Laboratory, Wright-Patterson AFB, Ohio, Technical Report AFML-TR-75-24, February 1975.

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT - PATTERSON AFB, OHIO
&
EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

ULTRASONIC TECHNIQUES FOR NONDESTRUCTIVE EVALUATION (NDE)

Prepared by:	M. Paul Hagelberg PhD.
Academic Rank:	Professor
Department and University	Department of Physics Wittenberg University
Assignment:	
(Laboratory)	Materials
(Division)	Metals and Ceramics
(Branch)	Nondestructive Evaluation
USAF Research Colleague:	Dennis Corbly
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

ULTRASONIC TECHNIQUES FOR NONDESTRUCTIVE EVALUATION (NDE)

By

M. Paul Hagelberg

ABSTRACT

Several ultrasonic techniques have been investigated for possible application to the problem of determining the quality of fit of a fastener and/or the existence of flaws in the hole wall. In the course of these investigations several gaps in the fundamental understanding of the underlying physical bases have been found. As a result, several lines of research have been suggested that are not only of interest in themselves but which may have considerable practical significance as well.

A study has been undertaken of the dependence of the velocity of the interface waves at the boundary between two solids on the frequency of the waves and the roughness of the surfaces. The materials used are aluminum, steel, and titanium. The effects are found to be substantial.

When an acoustic wave guide, such as a flat plate or a cylinder, is excited by a compressional pulse, the echo train can be extended by mode conversion at the boundaries. If the boundaries of the guide are altered, the mode conversion mechanism will be changed. By observing the echo train it may be possible to determine how the boundaries are altered. Preliminary observations have been made for the case of a flat plate, for a tapered pin in a thin plate, and for a Taper-Lok fastener holding a steel plate against an aluminum plate. Substantial changes can be seen when conditions at the boundaries are altered.

ACKNOWLEDGEMENTS

It is my pleasure to thank the American Society for Engineering Education and the Air Force Systems Command both for sponsoring this program and for inviting me to participate in it. I am grateful to the Materials Laboratory and Chief Scientist Frank Kelley for providing the facilities, equipment, and congenial surroundings in which I worked. May special thanks are extended to laboratory technicians David Dempsey and George Mescher for their willing and able assistance and to numerous other persons for help of one sort or another. Two persons deserve special thanks which I cannot adequately express. They are my Research Colleaque Dennis Corbly and Program Director J. Fred O'Brien, Jr. My work in this program was immeasurably aided by their tireless and able support.

ULTRASONIC TECHNIQUES FOR

NONDESTRUCTIVE EVALUATION (NDE)

By

M. Paul Hagelberg

INTRODUCTION

A serious limitation in a variety of technological areas is the lack of techniques for determining the integrity of such things as welds, castings, and mechanical joints and hence their ability to meet the specifications to which they are designed. A number of methods for nondestructive evaluation of structures are in use and although they are successful in many applications many problems remain.

One problem that is of particular concern in aircraft structures is that of inspecting fastener holes. A substantial number of structural failures result from fatique cracks that originate in such fastener holes. Two types of inspection can be considered. If the original fit between fastener and hole is a good one, then the probability of formation of a fatique crack is greatly reduced. Hence a technique for determining the quality of the fastener installation is desired. A second area of thrust is the development of methods for identifying and quantifying already existing cracks in structures that are in-service. It is important, especially from the point of view of cost, that this inspection be performed with the fastener in place. Ultrasonic methods seem to be the only promising approach that satisfies this latter criterion.

OBJECTIVES

The immediate objectives of the research undertaken during this program has been to isolate a small number of ultrasonic methods for nondestructive evaluation that hold promise of application to the fastener problem and to outline research procedures for establishing the physical bases for these methods. In approaching the research, consideration has been given to developing areas for study after this summer program has ended as well as to that which can be concluded during the period of the program. Two specific problems have been isolated for concentrated study. A third problem will be mentioned that has been defined but has not yet been given much attention because of the limited time available.

In the main body of this report each of the research areas will be described separately and the results of the investigations to date reported. Conclusions and recommendations will be consolidated into one section for all three areas.

INTERFACE WAVES

The propagation of guided elastic waves along a plane interface between two different elastic materials was first predicted by Stoneley (1). More recently Pilant (2) has determined the conditions under which the waves predicted by Stoneley can exist as well as conditions under which attenuated interface waves can propagate. It is predicted that interface waves will propagate at the boundary between such materials of engineering interest as aluminum and steel, aluminum and titanium, and titanium and steel. The object of this study is to verify the propagation of interface waves in these cases and to determine if they can be used to study the interface between an interference fit fastener and its hole. In addition, the effects of surface roughness on interface wave propagation is studies since such roughness is to be expected in real fasteners.

The experiment is conducted using a flat, polished plate of the more dense material to provide a free surface along which Surface Acoustic Waves (SAWs) are propagated. The travel time for the SAWs over a fixed path from transmitter to receiver is noted. The interface is obtained by pressing the polished face of the second material against the polished surface of the first so that the SAWs have no direct path from transmitter to receiver. Since interface waves travel at a slightly higher velocity than SAWs, a predictable decrease in travel time should be observed as the applied pressure produces a tight interface. Measurements made for aluminum on steel, aluminum on titanium, and titanium on steel give good agreement with the predicted interface wave velocities.

Since the motivation for this study is the probing of fastener-hole wall interfaces, the experiment has been extended to systems with axial symmetry. The surface waves are generated on a tapered steel pin which can be inserted in a tapered hole in a two-layer aluminum structure. The transit time for the surface wave to travel from the transducer to the end of the pin and back is measured. The pin is inserted in the plate with about 0.008" interference and the measurement is repeated. Not only is the change in travel time in agreement with the predicted value but one can see reflected waves from each surface in the structure as well. These measurements suggest that interface waves can provide useful information about the quality of an interference fastener fit.

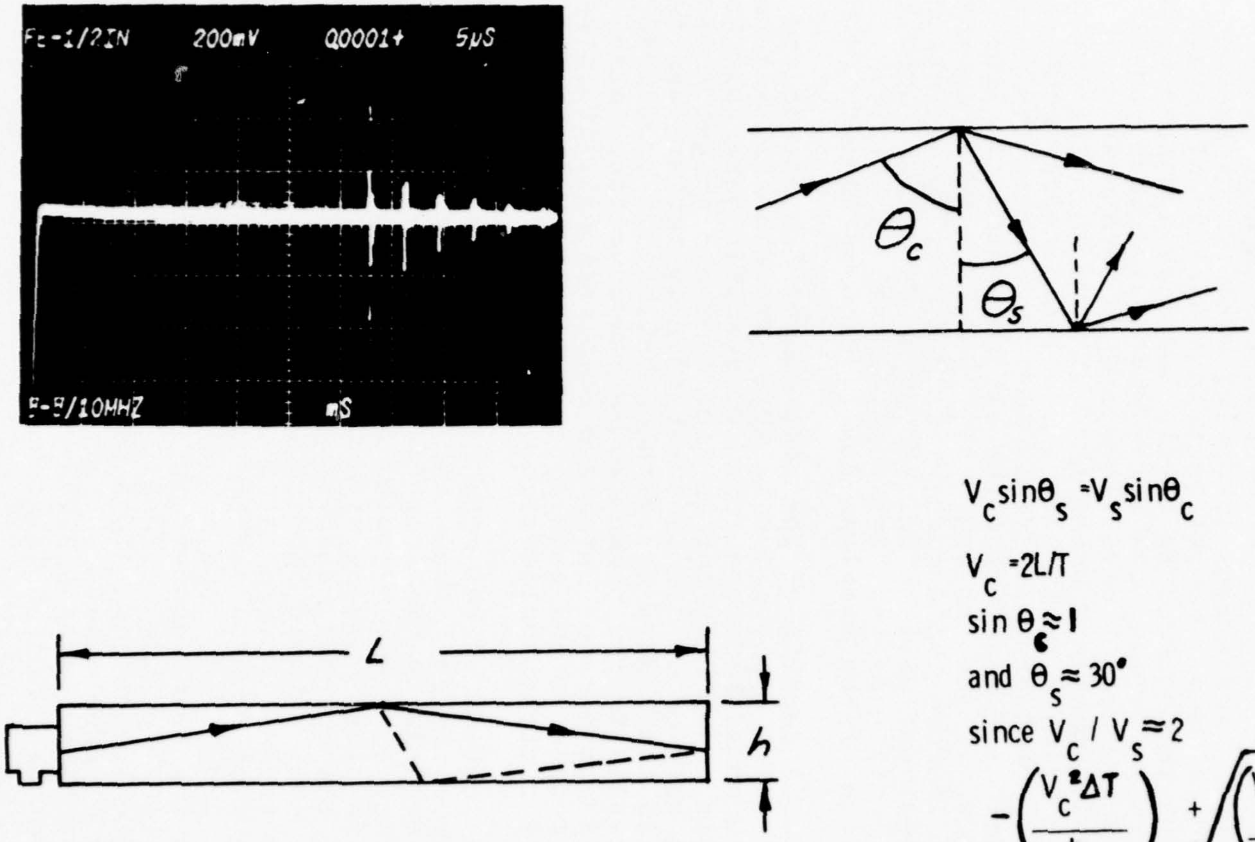
Work is being completed on a study of the effects of surface roughness on the propagation of interface waves. This is important because of the effects that may arise from the surface roughness left when parts are machined. To date it has been found that these effects are not insignificant and may be frequency dependent. These observed effects may influence, perhaps even enhance, the usefulness of interface waves in NDE applications.

Another related study that is in progress has to do with the reversal of roles of the two materials. If the SAW is originally generated in the lighter material, say the aluminum, the particle displacement field in the interface wave will be very different from that in the surface wave (3). As a result little energy will be converted at the boundary from free surface to interface. This invertigation is important as a check on the reliability of the theory. Our preliminary measurements indicate that the predictions of the theory are substantially correct.

MODE CONVERSION IN ACOUSTIC WAVEGUIDES

It has been known for some time (4) that when a compressional wave is generated in a sample such as a plate or cylinder with a dimension on the order of a few tens of wavelengths, the sample will act as an acoustic waveguide. An interesting feature of such a waveguide is that as the compressional wave energy is reflected from the guide walls, some is mode converted to shear wave energy. It then crosses the guide at a sharper angle and slower speed. When it reaches the other surface it is reflected again and may or may not be converted back to a compressional wave. The result is that a series of pulses are formed and the receiver pickes up the direct pulse, a second that is delayed by having crossed the guide once as a shear wave, a third that has crossed twice in shear form, and so on.

The photograph in Fig. 1 shows the echo train when a short compressional wave pulse is generated in a 0.500" steel plate 4.0" long. The secondary echoes corresponding to the mode converted waves are clearly visible. Below the



$V_c \sin\theta_s = V_s \sin\theta_c$

$V_c = 2L/T$

$\sin\theta_c \approx 1$

and $\theta_s \approx 30°$

since $V_c / V_s \approx 2$

$$V_s \approx \frac{-\left(\frac{V_c^2 \Delta T}{h}\right) + \sqrt{\left(\frac{V_c^2 \Delta T}{h}\right)^2 + \frac{4V_c^2}{\cos\theta_s}}}{2}$$

Figure 1

photograph is a sketch of the acoustic waveguide showing a possible path for the energy. This is shown expanded on the upper right.

Several possible uses for these guided and mode converted sound waves can be identified. One interesting application is as an experiment in the undergraduate physics laboratory. Experiments involving ultrasonics are uncommon yet would certainly be relevant in a variety of ways. This experiment can be used to obtain modestly accurate values for the compressional and shear wave velocities, to illustrate the concept of mode conversion, and to obtain values for the elastic constants using the measured sound velocities and the density. It is also quite inexpensive provided relatively standard electronic instruments are available. The equations for reducing the data are shown in Fig. 1. Since v_c/v_s 2 one can use $\cos\theta_s$ 0.8 as a first approximation and then improve the accuracy, if desired, by iteration. This experiment will be written up in detail and submitted for publication in the "American Journal of Physics", a publication of the American Association of Physics Teachers.

Several aspects of the mode conversion phenomenon have potential application in NDE. Two general features should be noted before turning to specific details. As can be seen in the photograph in Fig. 1, the signal-to-noise ratio is very good. It is substantially better than it appears in the figure where the baseline has broadened substantially in the photographic reproduction. Also, because the generated waves are compressional, it is very easy to couple to the sample, a particularly desirable feature for field applications.

Application of the mode conversion approach in NDE will result from the fact that the nature of the mode conversion process depends on the properties of the wave guide boundaries. If, for example, a second elastic medium were pressed in close contact with the surface of the guide material the boundary conditions would be substantially altered. This should show up as a substantial change in the echo train. Such alteration of the boundary conditions occur in such diverse situation as an interference fit fastener and for a plate adhesively bonded to another. Figure 2 shows the echo train and acoustic power spectrum for a free Taper-Lok fastener, on the left, and for the same fastener pressed into a hole in a structure consisting of a 0.400" steel plate and a 0.500" aluminum plate. The insertion distance is 0.222" providing about 0.0046" of interference. Substantial changes can be seen in both the

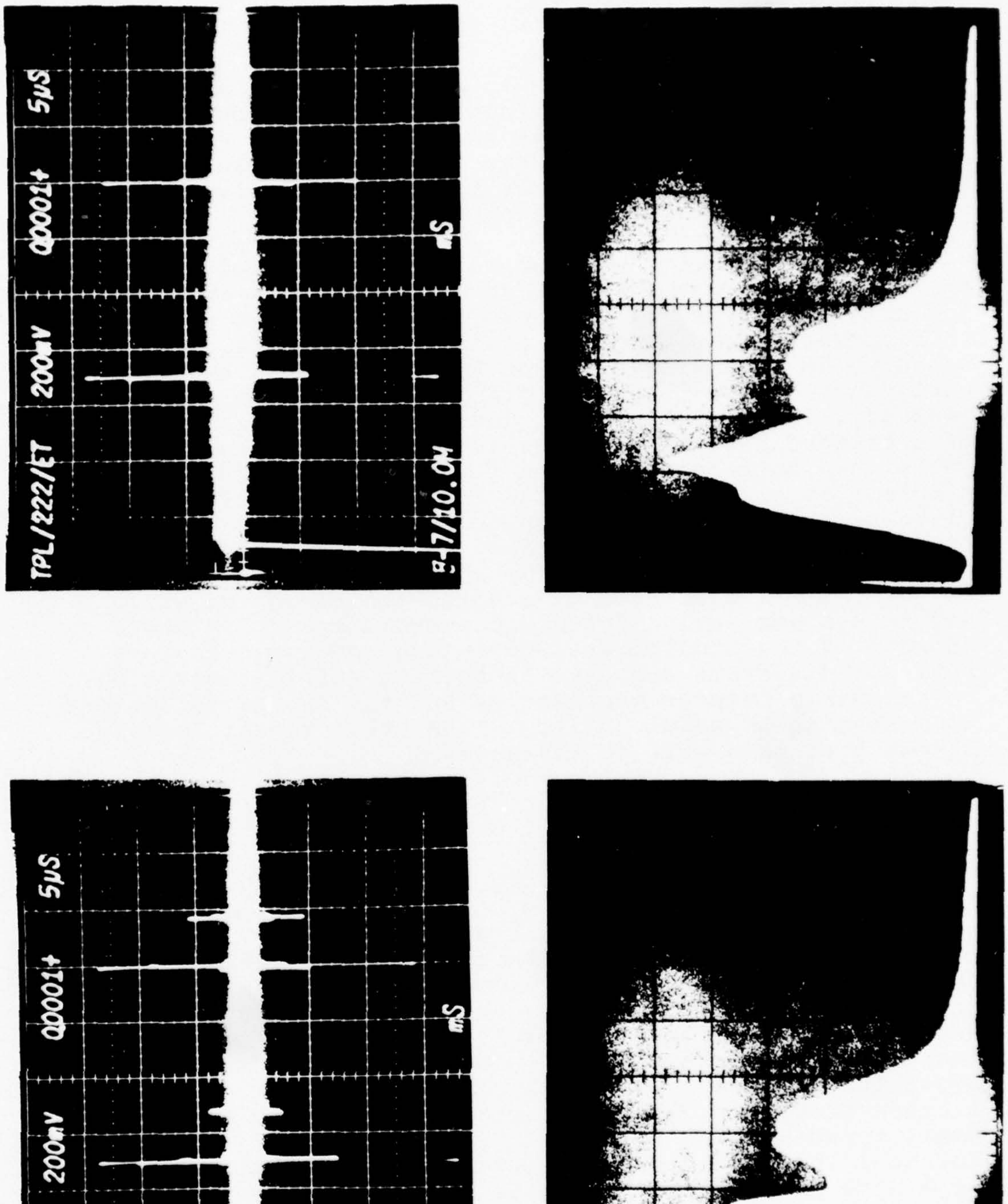

TPL/222/ET
200mV
0.0001+
5µS
mS
7/10.0M
TPL/FR/ECT
200mV
0.0001+
5µS
mS
7/10.0M

Figure 2

gross features of the echo train and in the acoustic power spectrum when the fastener is inserted. These changes should be related to the nature of the fit between hole and fastener and are potentially useful for the nondestructive evaluation of the fastener-hole interface.

Although Redwood's paper outlines the general features of the acoustic waveguide, a number of details must be investigated before one can use the technique effectively. This summer period has been too brief to permit extended research in this area. It has been shown that substantial variations can be observed both in systems with simple geometry, i.e., flat plates, and in the complex structure of a tapered fastener. A research program in this area is envisioned that begins with a study of the behavior of a simple flat plate to determine in detail the effects of plate thickness, length, transducer parameters, etc., on the acoustic energy propagation. After this aspect of the program has been completed the research will be extended to determine the effects of loading the waveguide walls and to systems having cylindrical symmetry. When these aspects of the problem are understood one can turn to real world systems and ask "Is there a sufficiently clear relationship between observables and the system parameters to make this a useful NDE tool?" At this time it seems likely that this will be the case.

SURFACE WAVES ON A CYLINDER

In the course of the study of waves on a tapered pin it was noted that in addition to those travelling along a generator of the surface, waves were observed to travel helically along the surface in certain well-defined paths. Time did not permit any additional study of these unexpected propagation properties. They do, however, seem worthy of additional investigation because their paths would permit inspection of certain surface features that a wave traveling along the generator would not be sensitive to. In particular, a crack growing from the interface would present a very small cross-section to a wave travelling along the generator but would have a component across the propagation direction of a wave propagating helically along the cylinder.

CONCLUSIONS AND RECOMMENDATIONS

Several ultrasonic techniques for the nondestructive evaluation of important problem configurations have been considered. The study of the general properties of interface waves has been essentially completed and it has been demonstrated that they can be used in fastener geometries. It is recommended that this work be evaluated to determine if it should be developed as a field method for evaluating interference fit fastener properties.

It has been shown that mode conversion in acoustic waveguides is sensitive to parameters that are related to fastener-hole quality and perhaps to other problem areas as well. A great deal is not presently known about this phenomenon. It seems appropriate that an extensive research program be initiated to determine the feasibility of its use in nondestructive evaluation.

The helical modes of propagation of surface waves on a cylinder hold promise of application to nondestructive evaluation of fasteners. Since it is a problem that can be pursued with modest equipment I intend to try to interest some undergraduates at Wittenberg in it. It should certainly be possible to do some of the preliminary work with no external funding. If it is shown to be a sufficiently promising approach, support can be sought at a later date.

REFERENCES

1. Stoneley, R., Elastic Waves at the Surface of Separation of Two Solids. Proc. Roy. Soc. (London), Ser. A 106, 416-428(1924).

2. Pilant, W. L., Complex Roots of the Stoneley - Wave Equation. Bull. Seism. Soc. Am. 62, 285-299(1972).

3. Lee, D. A., Private Communication.

4. Redwood, M., Velocity and Attenuation of a Narrow - Band, High Frequency Compressional Pulse in a Solid Wave Guide. J. Acoust. Soc. Am. 31, 442-448(1959).

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT-PATTERSON AFB, OHIO

&

EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

FRACTURE OF GRAPHITE/EPOXY COMPOSITES

Prepared by:	Don H. Morris, Ph. D.
Academic Rank:	Associate Professor
Department and University:	Mechanical Engineering Department Mississippi State University
Assignment:	
(Laboratory)	Materials
(Division)	Nonmetallic Materials
(Branch)	Mechanics and Surface Interactions
USAF Research Colleague:	Dr. H. T. Hahn
Date:	August 15, 1975
Contract No.:	F44620-75-C-0031

FRACTURE OF GRAPHITE/EPOXY COMPOSITES

By

Don H. Morris

ABSTRACT

The resistance method is applied to composite materials. Experimental data for center cracked tensile specimens reveals a linear relationship between crack growth resistance and initial crack length. Comparisons are made between laminates of the same ply orientation but of different thicknesses, and between laminates of different ply orientations.

There is evidence that the resistance method, for one of the laminates, reduces to the inherent crack model. More complete experimental data is needed to fully verify this statement.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the support of the USAF and ASEE for providing a fellowship through the Summer Faculty Research Program. Special thanks is given to Dr. H. T. Hahn for providing assistance throughout the period of this study. The author also acknowledges Dr. R. Kim for testing the composite material specimens. Thanks is also due to Dr. J. M. Whitney for his support of the program, and for allowing the author to use equipment in his laboratory.

INTRODUCTION

The load-carrying capacity of a structural member is greatly reduced by the presence of cracks or crack-like defects. Thus, a significant effort has been directed toward the prediction of loads which cause catastrophic failure in structural components. Consequently, several methods exist for predicting the strength of flawed isotropic materials, using the method of linear elastic fracture mechanics (LEFM).

LEFM has also been successfully applied to several resin-matrix composite material systems. Waddoups, Eisenmann, and Kaminski [1] used an inherent flaw concept to explain the hole size effect in tension coupons containing circular holes. Konish, Swedlow, and Cruse [2] used limited experimental data to show that the critical value of the stress intensity factor and the critical energy release rate at fracture are material constants for a given laminate. They used LEFM to reduce data for a graphite/epoxy material. Cruse [3] was able to predict the fracture strength of a laminate based on its constituent angle ply fracture data. He also presents a macromechanics explanation of the notch size effect on static strength for circular holes.

Whitney and Nuismer [4], using two related criteria based on normal stress distribution, obtained a two parameter (unnotched tensile strength and a characteristic dimension) model which was capable of predicting discontinuity size effects without application of LEFM. However, both criteria lead to a relationship between Mode I fracture toughness and unnotched laminate tensile strength.

Tsai and Hahn [5] mention the possibility of using the resistance method [6] of characterizing the stable growth of crack damage in the region of a crack tip. The resistance method has been applied to plane-stress fracture toughness evaluation of metals [7]; its extension to anisotropic composite materials is as yet unknown. The work reported herein will consider the application of the resistance method to graphite/epoxy composites.

OBJECTIVE AND SCOPE

The objective of this investigation was to consider the possibility of using the resistance method as a means of predicting catastrophic failure of graphite/epoxy composites. Crack growth resistance curves (K_R-curves) were obtained by testing center-cracked tension specimens of $[0/90/\pm45]_{2s}$, $[0/\pm45]_{2s}$, and $[0/\pm45]_s$ graphite/epoxy laminates, with center-cracks ranging from 0.2 in. to 1.0 in. K_R-curves were generated from crack length-crack opening displacement data. Results allow a comparison between different laminates $[0/90/\pm45]$ and $[0/\pm45]$ and different thicknesses of the same laminate $[0/\pm45]_s$ and $[0/\pm45]_{2s}$.

RESISTANCE METHOD

Under conditions of plane-strain, high-strength metals do not exhibit slow, stable crack growth prior to unstable fracture. However, thin metal sheets do exhibit slow stable crack growth prior to catastrophic failure. In addition, fracture toughness (K) or strain energy release rate (G) values are dependent upon specimen width and initial crack size. For these reasons, i.e., lack of a unique value of K for a particular material, and stable crack growth, an attempt is being made to apply the resistance method to the plane-stress fracture toughness evaluation of metals [7]. In fact, Krafft, Sullivan, and Boyle [8] postulated that for a given material and thickness there is a unique relationship between the amount a crack grows and the applied stress intensity factor. This relationship is called a crack growth resistance curve (hereafter called a K_R-curve). As noted by Sullivan, Freed, and Stoop (in Ref. 7), a geometry independent K_R-curve can be a useful tool to engineers engaged in failure-safe design.

A great deal of effort has been expended on K_R-curve determination of metals [7]. However, no attempt has been made to extend the method to composite materials. Prior to determining K_R-curves for graphite/epoxy laminates, a brief discussion is given on crack growth resistance curves. Further details may be found in Ref. 7.

Upon increasing the load, an increase occurs in the rate of energy available for crack extension (G). This energy is opposed by an increasing resistance to crack extension (K_R). G and K_R remain in equilibrium up to the point of instability (point c in Fig. 1); this point is used to compute the stress intensity at commencement of unstable crack propagation. The G curves in Fig. 1 are plotted as a function of crack length (using the equation for K for the particular Mode I specimen being tested), where each curve represents a particular value of applied stress as a parameter.

Unstable crack extension occurs where the G and K_R curves are tangent (point c); the point of tangency is found from

$$G(a, \sigma_c) = K_R\ (a, \sigma_c) \qquad (1)$$

and

$$\left(\frac{\partial G}{\partial a}\right)_{\sigma_c} = \left(\frac{\partial K_R}{\partial a}\right)_{\sigma_c} \qquad (2)$$

These equations determine the fracture toughness and the critical crack length. A structure is then designed such that the stress-crack length relationship that may occur in service is less than the critical relationship.

In testing composite materials it was found that there is no visible slow, stable crack growth. Here, an effective crack length may be defined by matching the compliance based on the crack opening displacement [5]. The matching method is discussed in detail in the following sections, along with a discussion of the resistance curves obtained by the matching technique.

EXPERIMENTAL PROGRAM

The experimental program was designed to determine:

1. the effect of laminate orientation, crack length, and laminate thickness on the Mode I fracture toughness, and to compare the results with available experimental data (this is a secondary goal).

2. the nature of resistance curves of graphite/epoxy laminates, and to compare the results with other methods of predicting ultimate failure (this is the primary goal of the program).

A total of thirty-five center-cracked tension specimens were tested. All specimens were 2-in. wide, 12-in. long (9-in. between end tabs), with crack lengths 0.2, 0.4, 0.6, 0.8, and 1.0 in. The material was supplied by Whittaker Corporation, and consisted of Thornel 300 graphite fibers in a Narmco 5208 epoxy resin. Cracks were produced by first drilling a small hole in a specimen, followed by a final lengthening with a 5 mil diamond wire. No attempt was made to further sharpen the crack tips. A schematic of a test specimen is shown in Fig. 2; an actual specimen revealing a crack is shown in Figs. 3 and 4.

The laminate orientations and number of specimens may be summarized as:

$[0/90/\pm45]_s$; two specimens of each crack length

$[0/\pm45]_s$; two specimens of each crack length

$[0/\pm45]_{2s}$; three specimens of each crack length

The unnotched tensile strength was determined by testing six tensile coupons of the $[0/90/\pm45]_s$ laminate and four coupons of the $[0/\pm45]_{2s}$ laminate.

All specimens were loaded by friction grips, and tested in a closed loop MTS machine at a constant cross-head rate of 0.04 in/min. During each test, the applied load and crack opening displacement (COD) were continuously monitored and recorded. COD was measured by a double cantilever clip gage of the type used in fracture testing of metals [9]. The clip gage was attached to aluminum tabs which were bonded to the specimen using epoxy cement. M-Bond 200 strain gage cement was found to be unsuitable due to debonding between tabs and specimen before specimen fracture. An attached clip gage may be seen in Fig. 4. Calibration of the clip gage was performed in an Instron extensometer calibrator.

A comparison of three test records of load-COD is seen in Fig. 5, where each specimen had the same initial crack length. These records indicate rapid changes in COD, similar to the pop-in effect seen in metals. A failed specimen is seen in Fig. 6 ($[0/\pm45]_{2s}$, crack length = 0.6 in.). Figure 7 shows the same specimen separated to reveal delamination and fiber fracture. A comparison of the fracture of two different laminates is shown in Fig. 8. At this time, no statement is made regarding the difference in fracture surfaces.

Nominal stress at fracture and fracture toughness were calculated using the fracture load read from the load-COD records. The relationship between nominal stress at fracture and initial crack length is shown in Fig. 9, the values indicated being the average of two or three specimens, depending on the particular laminate. The only significant difference between different laminates occurs when there is no crack. The values indicated in Fig. 9 for the $[0/90/\pm45]_{2s}$ laminate are significantly less than those given by Nuismer and Whitney [10] for $[0/\pm45/90]_{2s}$ laminate, whereas the $[0/\pm45]_{2s}$ values compare favorably with those of Cruse and Osias [11].

The candidate fracture toughness [2] was calculated using the equation [9],

$$K = Y \sigma \sqrt{a} \tag{3}$$

where:

σ = nominal stress at fracture load
a = original half-crack length
Y = isotropic finite correction factor

The use of the isotropic finite width correction factor for anisotropic materials has been validated by Snyder and Cruse [12]. The equation for Y is given in [9],

$$Y = 1.77 \left[1 - 0.1 \left(\frac{2a}{W} \right) + \left(\frac{2a}{W} \right)^2 \right] \tag{4}$$

The candidate fracture toughness, original crack length relationship is shown in Fig. 10. The laminate values are less, in both cases, than those given in [10] and [11]. Both the variation in nominal fracture stress and candidate fracture toughness may be due, in part, to smaller unnotched tensile strengths than those cited in [10] and [11].

Figures 9 and 10 illustrate little, if any, effect of laminate thickness for the $[0/\pm45]_s$ and $[0/\pm45]_{2s}$ specimens. The fracture toughness of the $[0/90/\pm45]_s$ laminate is somewhat greater than either of the two $[0/\pm45]$ laminates, whereas the unnotched tensile strength of the two different laminates shows the reverse trend.

RESISTANCE CURVES

In order to construct resistance curves, it was first necessary to obtain a calibration between COD and crack length. As previously mentioned, COD was measured with a clip gage, as shown in Fig. 4. The gage length of the clip gage is the same for all specimens, approximately 0.3 in. From the initial straight portion of the load-COD record, the ratio of COD to load was determined, and plotted as a function of initial crack length. Calibration curves for converting COD measurements to crack length are shown in Fig. 11. These curves obviously depend upon type of laminate and laminate thickness.

Next, straight lines are drawn from the origin of the load-COD curve at increasing values of COD, and the inverse slope, COD/P, is determined. These values are used, together with the calibration curve, to determine effective crack lengths. As noted in [5], the effective crack length is not a pre-existing crack, but rather it is a crack-like region developing prior to the commencement of ultimate fail re. In fact, observation of a test revealed no actual crack length extension prior to catastrophic failure.

Crack growth resistance, K_R, is calculated from

$$K_R = Y\sigma\sqrt{a} \qquad (5)$$

where σ is the nominal stress based on various load levels and unnotched area, a is the effective crack length from calibration curves, and Y is the isotropic correction factor (eqn.4) based on effective crack length. K_R-curves for the three laminates tested are shown in Figs. 12, 13, and 14.

These figures illustrate several interesting features. The relationship between crack growth resistance and effective crack half-length, for both types of laminates, can be represented by a straight line. It should be noted that the straight lines are "eye-ball" fits; no attempt was made to use curve fitting techniques.

The maximum value of K_R is not constant, except possibly for the $[0/\pm45]_{2s}$ laminate. As previously stated, three specimens of each crack length were tested for this laminate, whereas the other two laminates tested consisted of only two specimens for each crack length. Due to data scatter in composite material testing, more experiments should be performed to determine whether or not the maximum value of K_R is constant. In addition, the values of K_R where load-COD curves deviate from a linear relationship are not constant.

Slopes of the K_R-crack half-length curves are not the same for each initial crack length, for a given type of laminate. However, the curves for different thicknesses of the [0/±45] material are approximately equal, indicating that laminate thickness may not be a variable in K_R testing. Once again, more data is needed to substantiate this statement. Only one data point, at the linear limit, is shown for the $[0/90/\pm45]_s$ laminate, since the load-COD response was linear to fracture.

In order to determine the critical values of K_R or crack length, it is necessary to locate the point of tangency between K_R and G, as shown in Fig. 1, and given by eqn. (2). Using eqn. (3) and calculating K_R (or G) as a function of crack half-length for various value of stress, it can be shown that there is no point of tangency. Thus, the critical values of fracture toughness and crack growth are the maximum values on the curves.

Another interesting feature is seen in Fig. 15, where maximum crack growth ($\Delta a = a - a_o$) is plotted against initial crack half-length. The horizontal lines represent average values for all crack lengths. The average values of Δa for two thicknesses of [0/±45] laminate are essentially the same, while the value for the [0/90/±45] laminate is much less. Again, note the data scatter. The horizontal line represents the data more favorably for three tests at each crack length than for two tests, again indicating the need for testing a large number of samples in composite evaluation.

Crack growth, Δa, is apparently independent of initial crack length for the [0/±45] laminate. As noted by Tsai and Hahn [5], the resistance method then reduces to the inherent crack model [1].

CONCLUSIONS AND RECOMMENDATIONS

The resistance method has been applied to composite materials, and the relationship between crack growth resistance and effective crack length was found to be linear. The critical values of fracture toughness and crack length are given by the maximum values on the K_R-curves.

Since construction of K_R-curves is dependent on compliance calibration, it is imperative that compliance curves be accurately determined. In future work, it is recommended that more than two or three tests be conducted, not only for compliance calibration, but for more accurate calculations of other quantities as well.

It was found that crack growth is apparently independent of initial crack length and laminate thickness for the [0/±45] material. The resistance method then reduces to the inherent crack model.

When compared to the efforts of other investigators, favorable agreement was found for nominal stress at fracture for $[0/\pm45]_{2s}$ laminate. However, the fracture toughness at fracture is less than that reported by two previous researchers; the difference may be due to lower unnotched strengths in the present investigation. In addition, there is little, if any, effect on laminate thickness for the [0/±45] material.

Future efforts in composite materials testing should be directed toward testing large quantities of specimens due to data scatter. Additional work along the line of the present investigation would be interesting, especially to determine if crack growth and maximum fracture toughness are actually independent of initial crack length. A somewhat different, but related problem, would be to determine the relationship between different modes of fracture when a crack is not perpendicular to the applied load. This, together with the present and previous work, would enhance the techniques of evaluating structural loads.

REFERENCES

1. Waddoups, M. E., Eisenmann, J. R. and Kaminski, B. E., "Macroscopic Fracture Mechanics of Advanced Composite Materials", J. of Composite Materials, pp. 446-454, Vol. 5 (Oct. 1971).

2. Konish, H. J., Jr., Swedlow, J. L. and Cruse, T. A., "Experimental Investigation of Fracture in an Advanced Fiber Composite", J. of Composite Materials, pp. 114-124, Vol. 6 (Jan. 1972).

3. Cruse, T. A., "Tensile Strength of Notched Composites", J. of Composite Materials, pp. 218-229, Vol. 7 (April 1973).

4. Whitney, J. M. and Nuismer, R. J., "Stress Fracture Criteria for Laminated Composites Containing Stress Concentrations", J. of Composite Materials, pp. 253-265, Vol. 8 (July 1974).

5. Tsai, S. W. and Hahn, H. T., "Failure Analysis of Composite Materials", to be presented at the Symposium on Inelastic Behavior of Composite Materials, ASME 1975 Winter Annual Meeting, Houston, Texas, Nov. 30-Dec. 5, 1975.

6. Bluhm, J., "Resistance Method", in Fracture Mechanics of Aircraft Structures, H. Liebowitz, ed., AGARD-AG-176, North Atlantic Treaty Organization, 1974.

7. Fracture Toughness Evaluation by R-Curve Methods, American Society for Testing and Materials STP 527, (1973).

8. Krafft, J. M., Sullivan, A. M. and Boyle, R. W., "Effect of Dimensions on Fast Fracture Instability of Notched Sheets", in Proceedings, Crack Propagation Symposium, College of Aeronautics, Cranfield, England, Vol. 1, pp. 8-28 (1961).

9. Brown, W. F. and Srawley, J. E., Plane Strain Crack Toughness Testing of High Strength Metallic Materials, ASTM STP 410, p. 144 (1966).

10. Nuismer, R. J. and Whitney, J. M., "Uniaxial Failure of Composite Laminates Containing Stress Concentrations", presented at ASTM Conference on Fracture Mechanics in Composites, Gaithersburg, Maryland, Sept. 25, 1974.

11. Cruse, T. A. and Osias, J. R., "Exploratory Development on Fracture Mechanics of Composite Materials", AFML-TR-74-111 (April 1974).

12. Snyder, M. D. and Cruse, T. A., "Crack Tip Stress Intensity Factors in Finite Anisotropic Plates", AFML-TR-73-209 (August 1973).

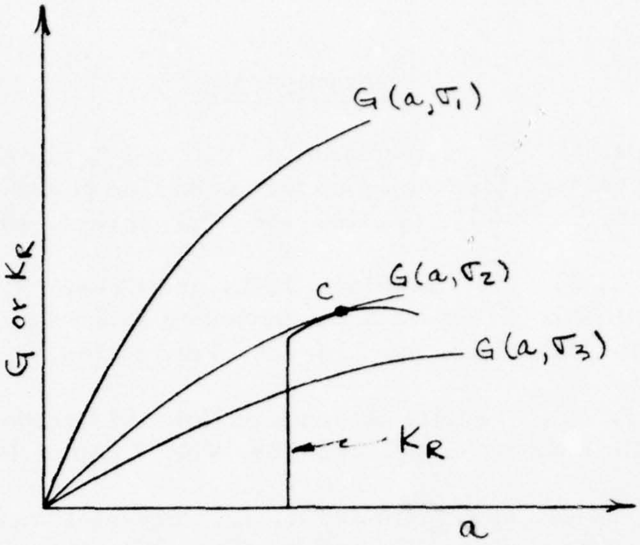


Fig. 1. Schematic of resistance method

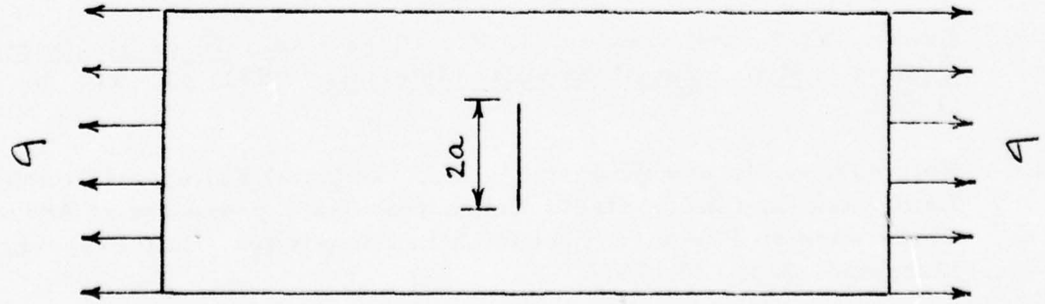


Fig. 2. Schematic of test specimen

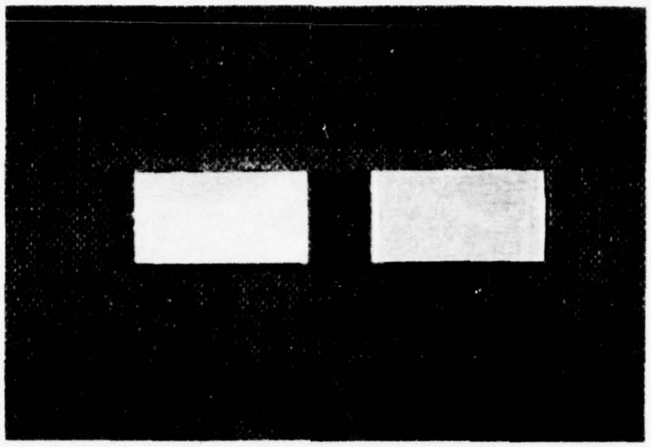


Fig. 3. Center-cracked specimen

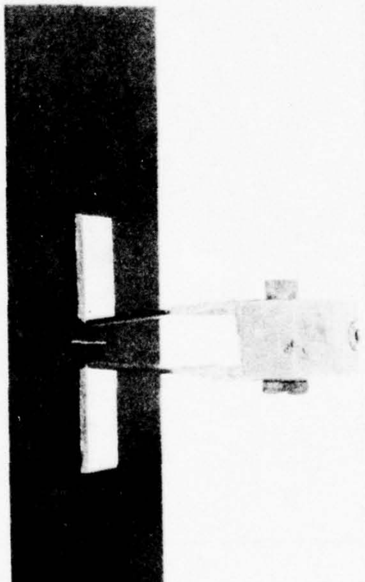


Fig. 4. Tensile specimen with clip gage attached

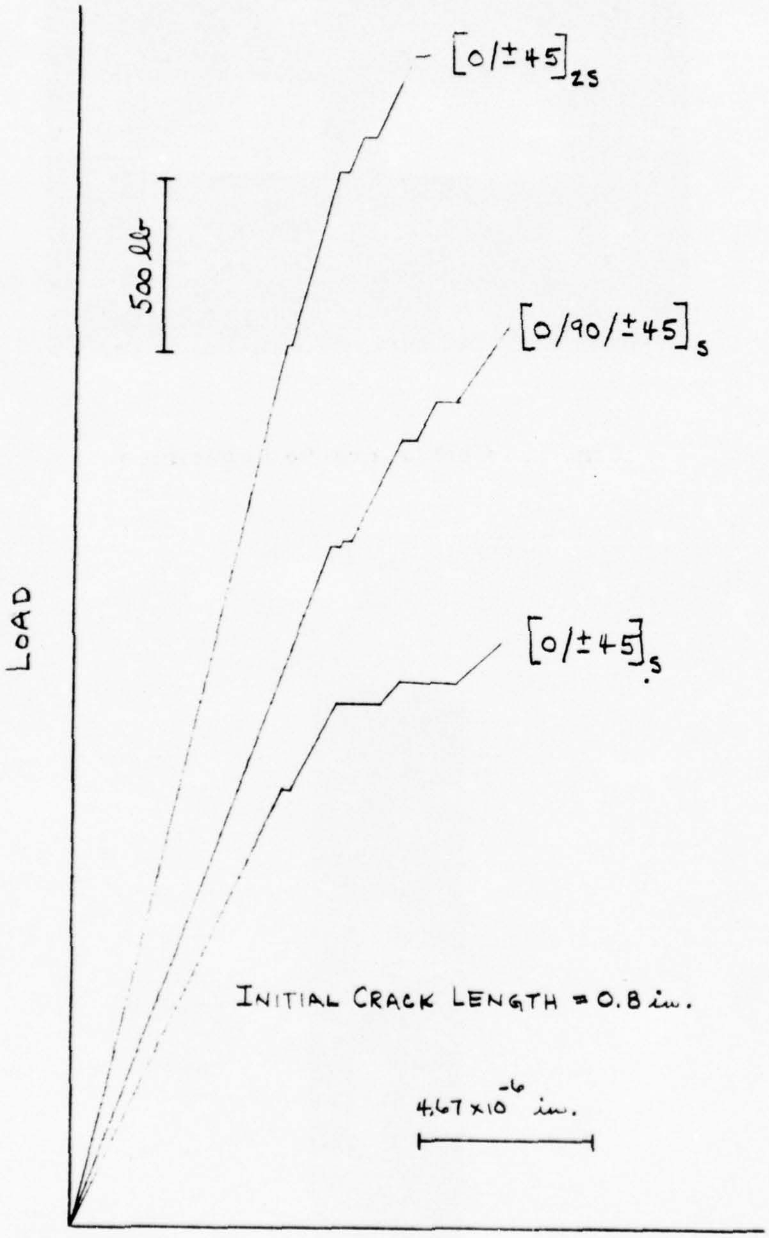


Fig. 5. Comparison of load-crack opening displacement records

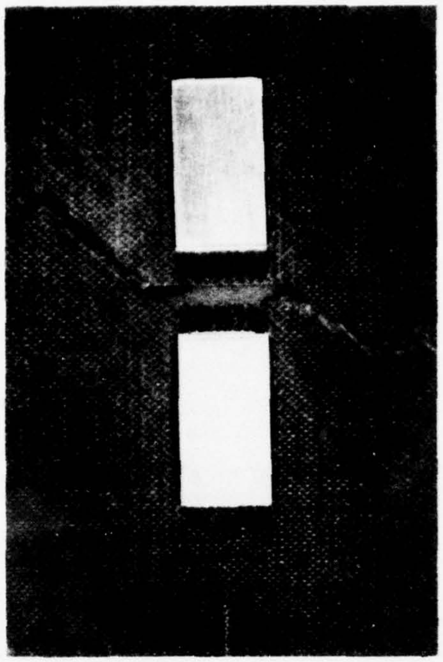


Fig. 6. As fractured specimen, $[0/\pm 45]_{2s}$

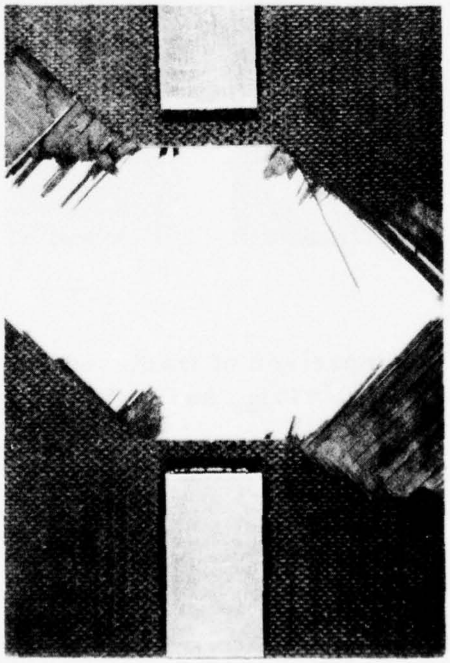


Fig. 7. Separated to reveal delamination and fiber fracture, $[0/\pm 45]_{2s}$

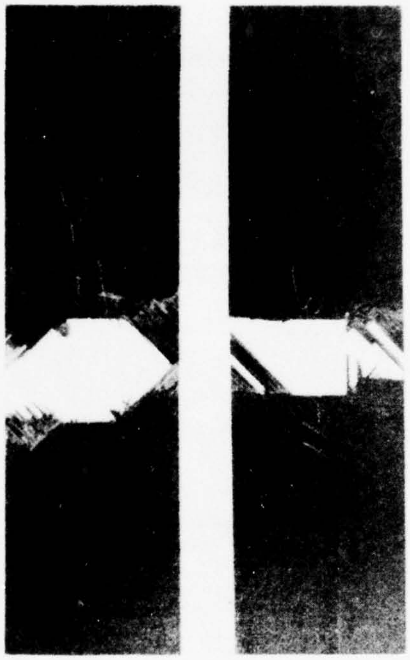


Fig. 8. Comparison of fractured specimens: left $[0/\pm45]_{2s}$ and right $[0/90/\pm45]_s$

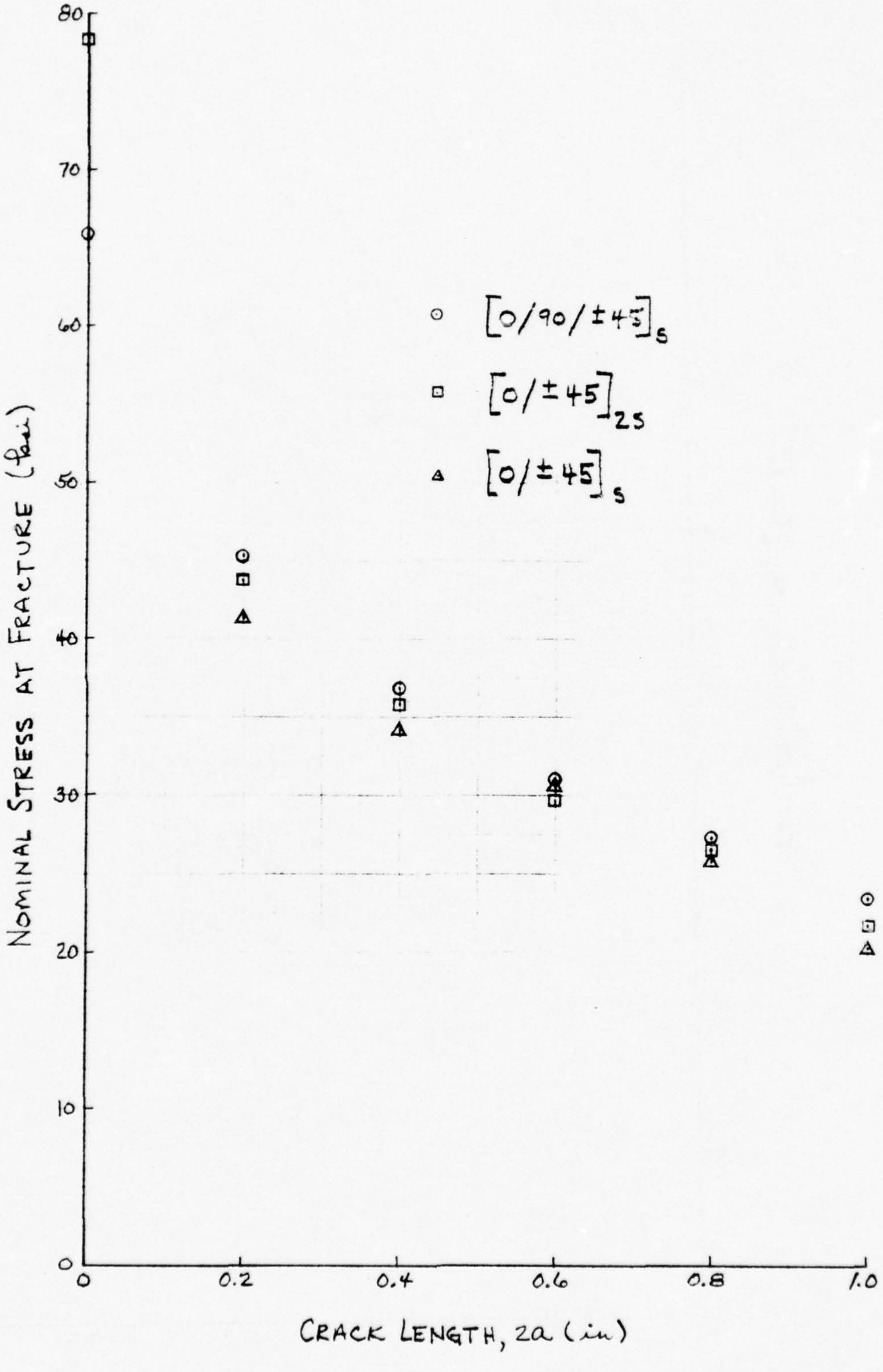


Fig. 9. Nominal stress at fracture vs. initial crack length

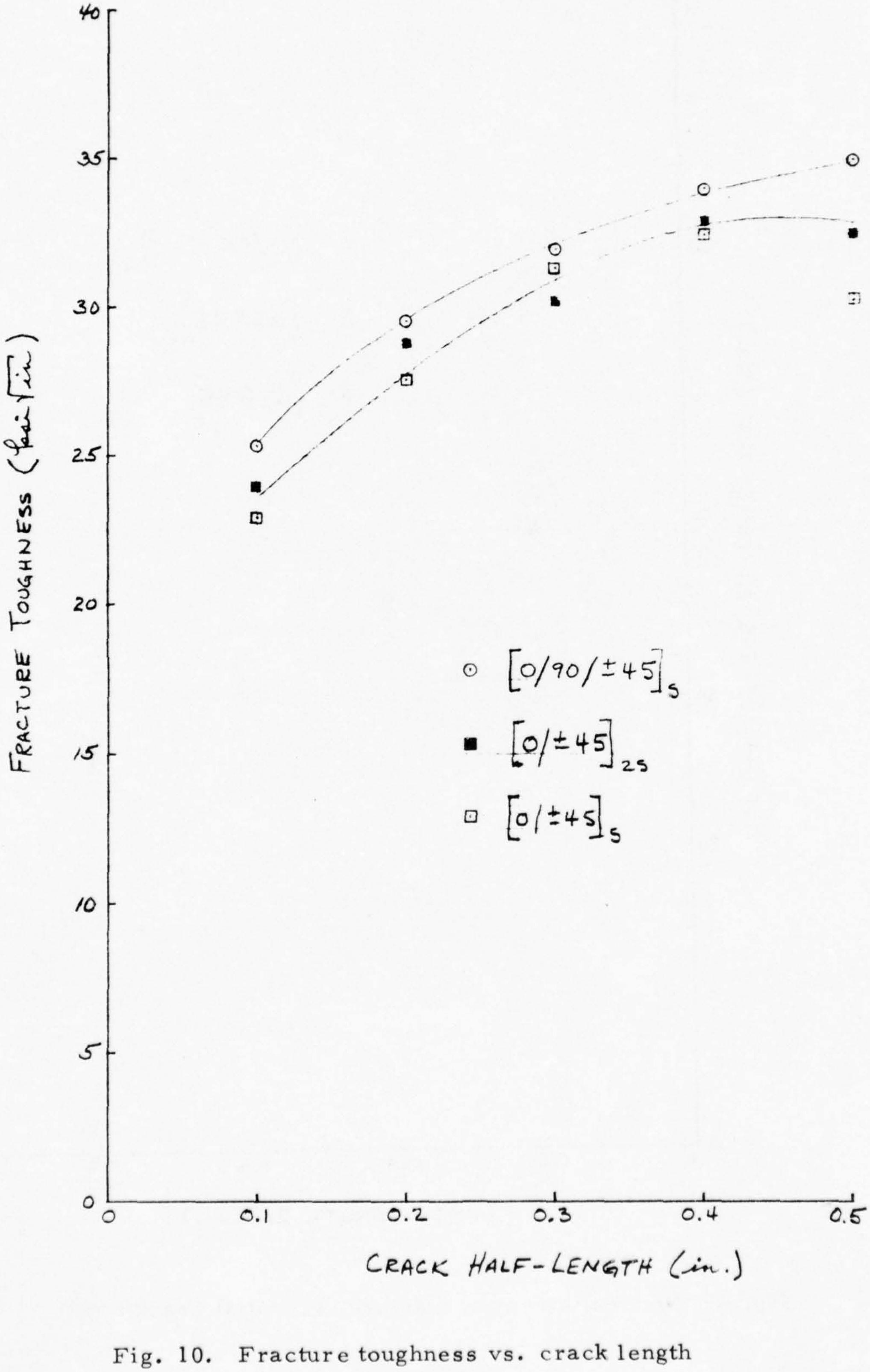


Fig. 10. Fracture toughness vs. crack length

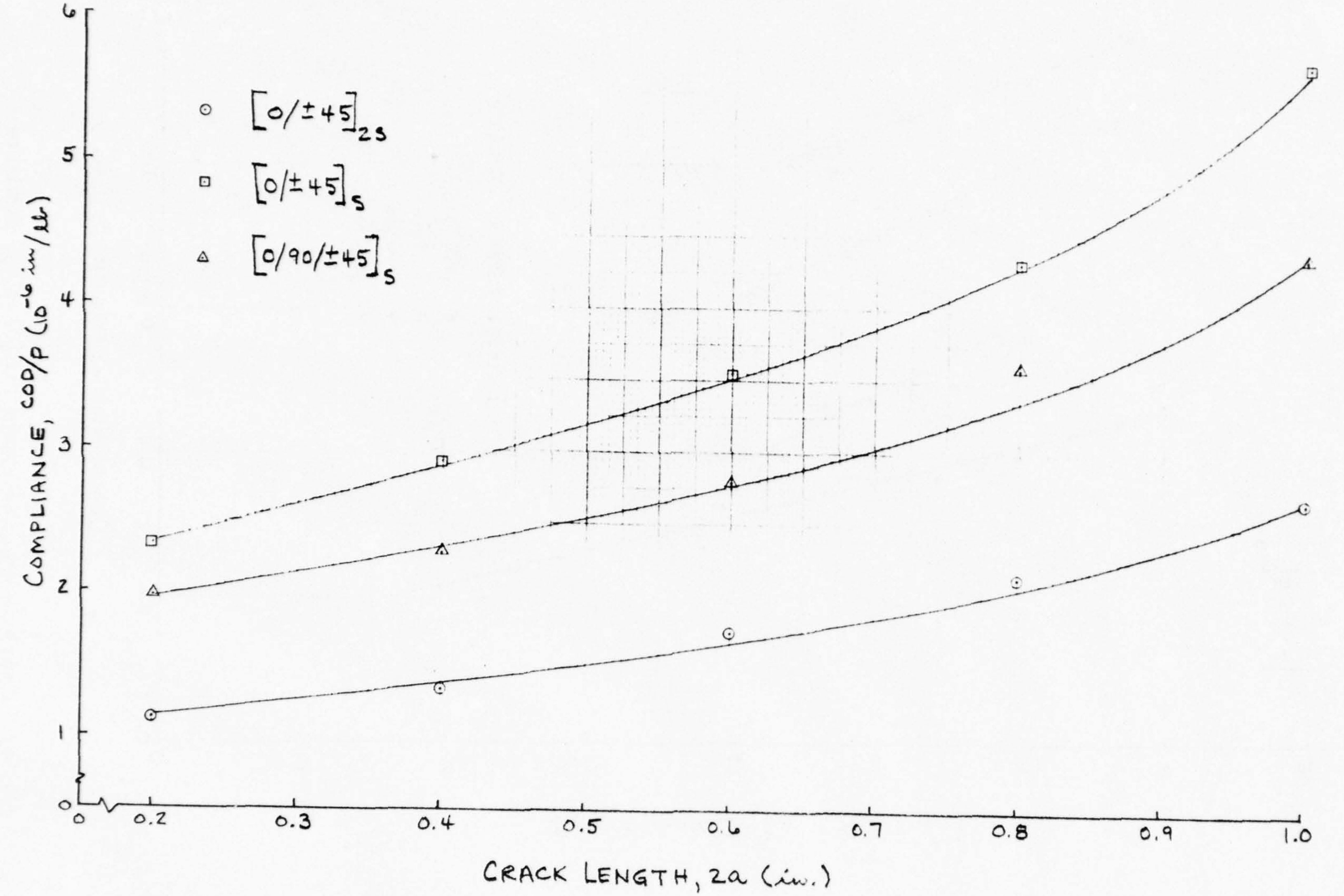


Fig. 11. Compliance curves for three laminates

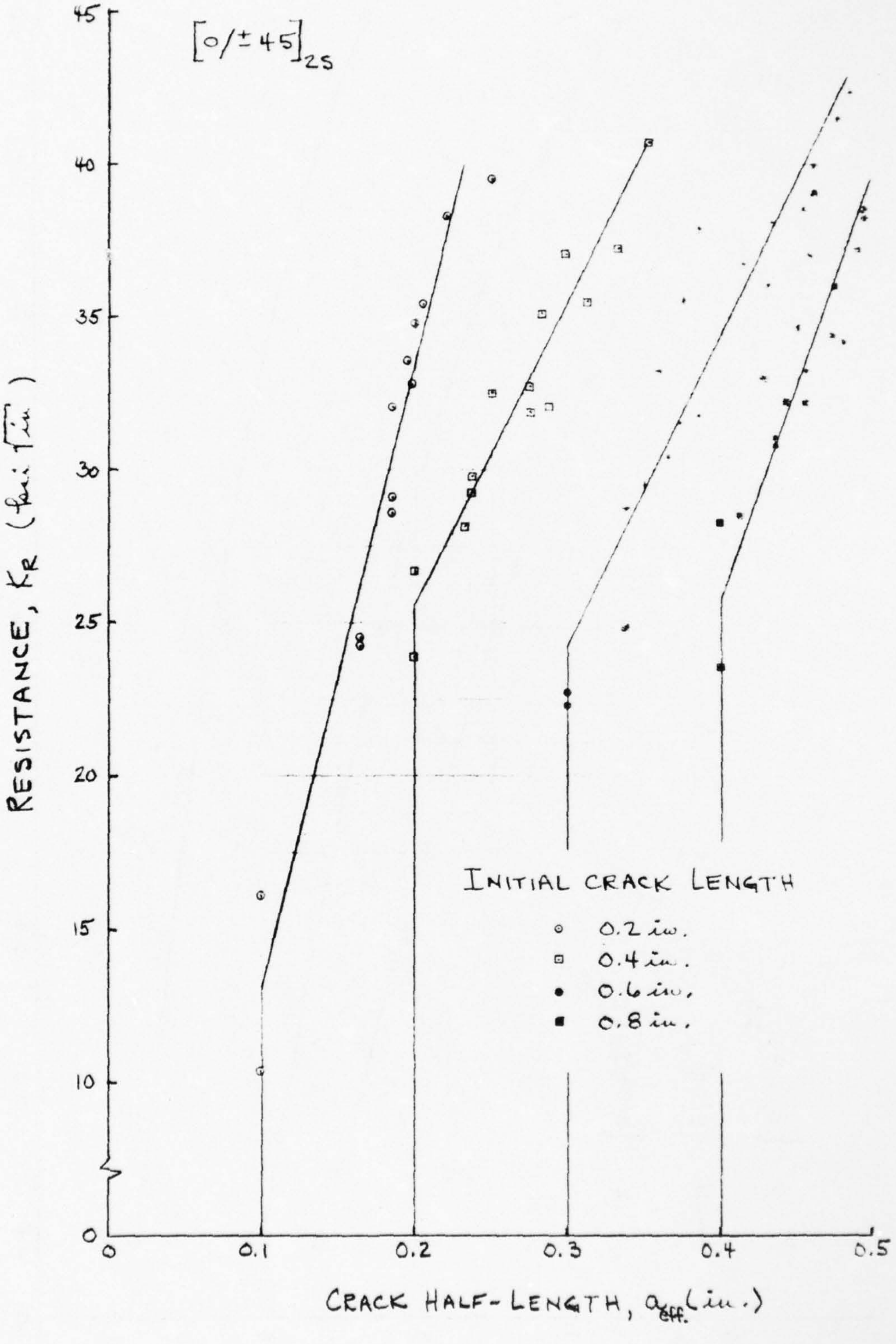


Fig. 12. Crack growth resistance for $[0/\pm45]_{2s}$ laminate

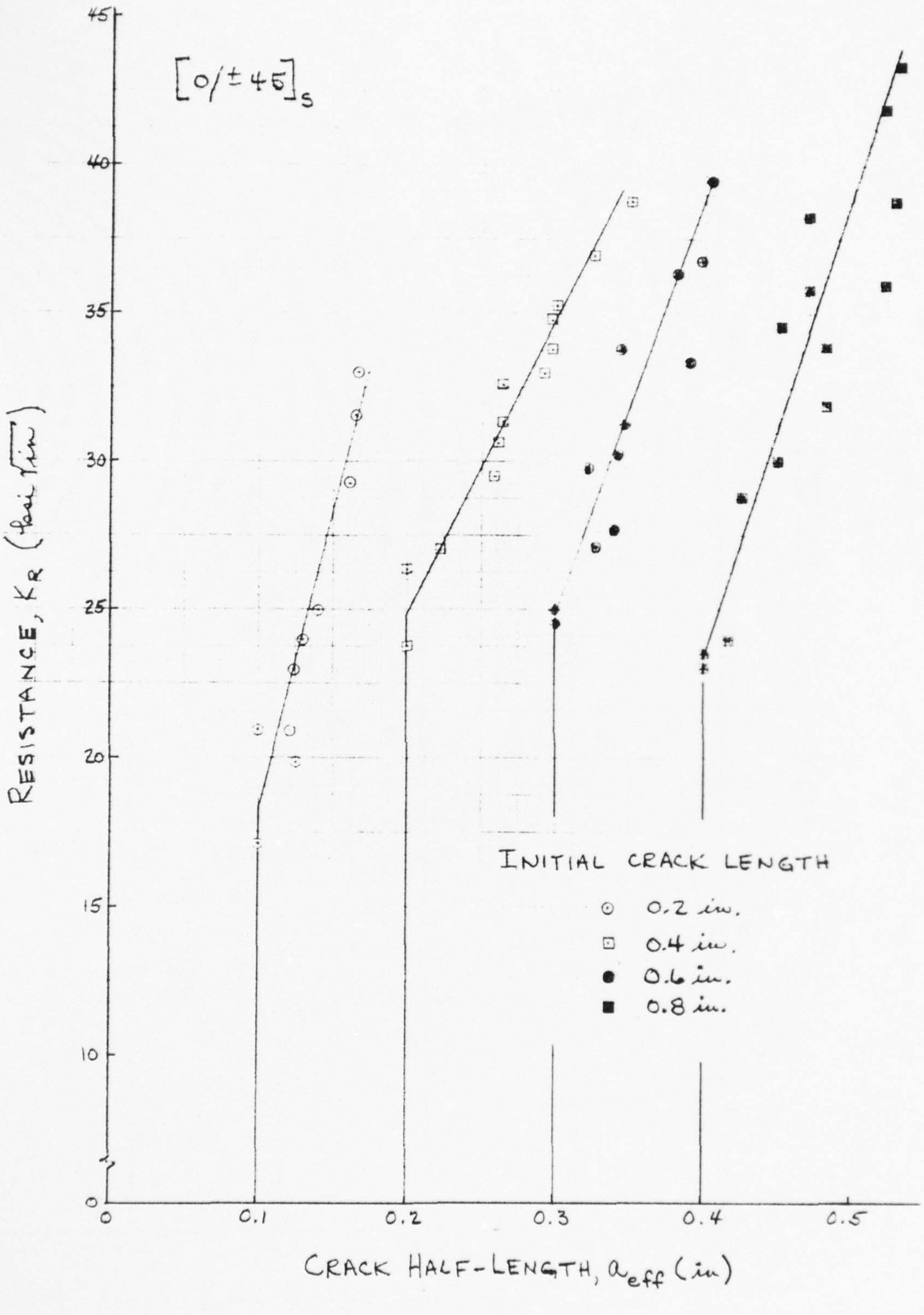


Fig. 13. Crack growth resistance for $[0/\pm45]_s$ laminate

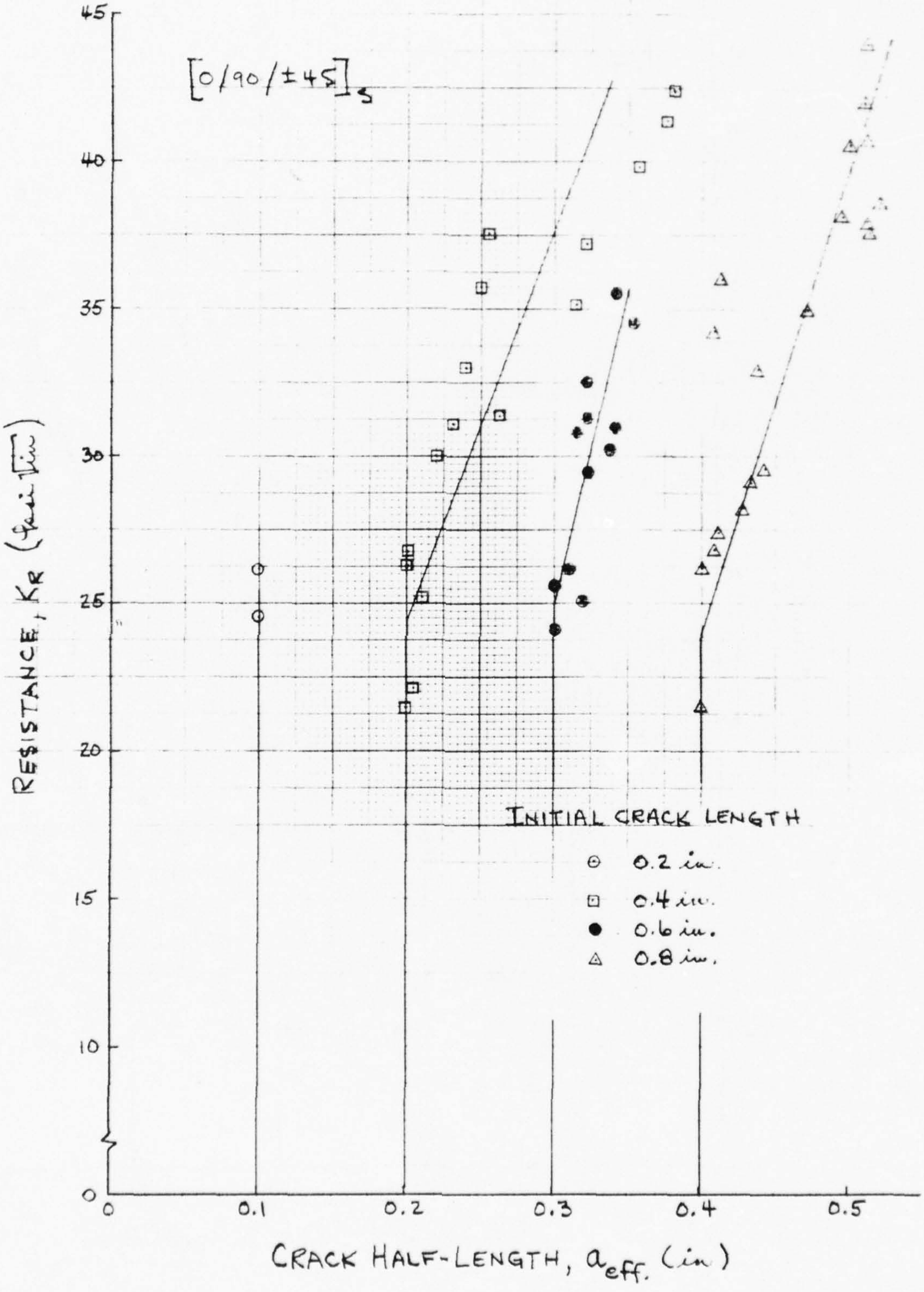


Fig. 14. Crack growth resistance for $[0/90/\pm45]_s$ laminate

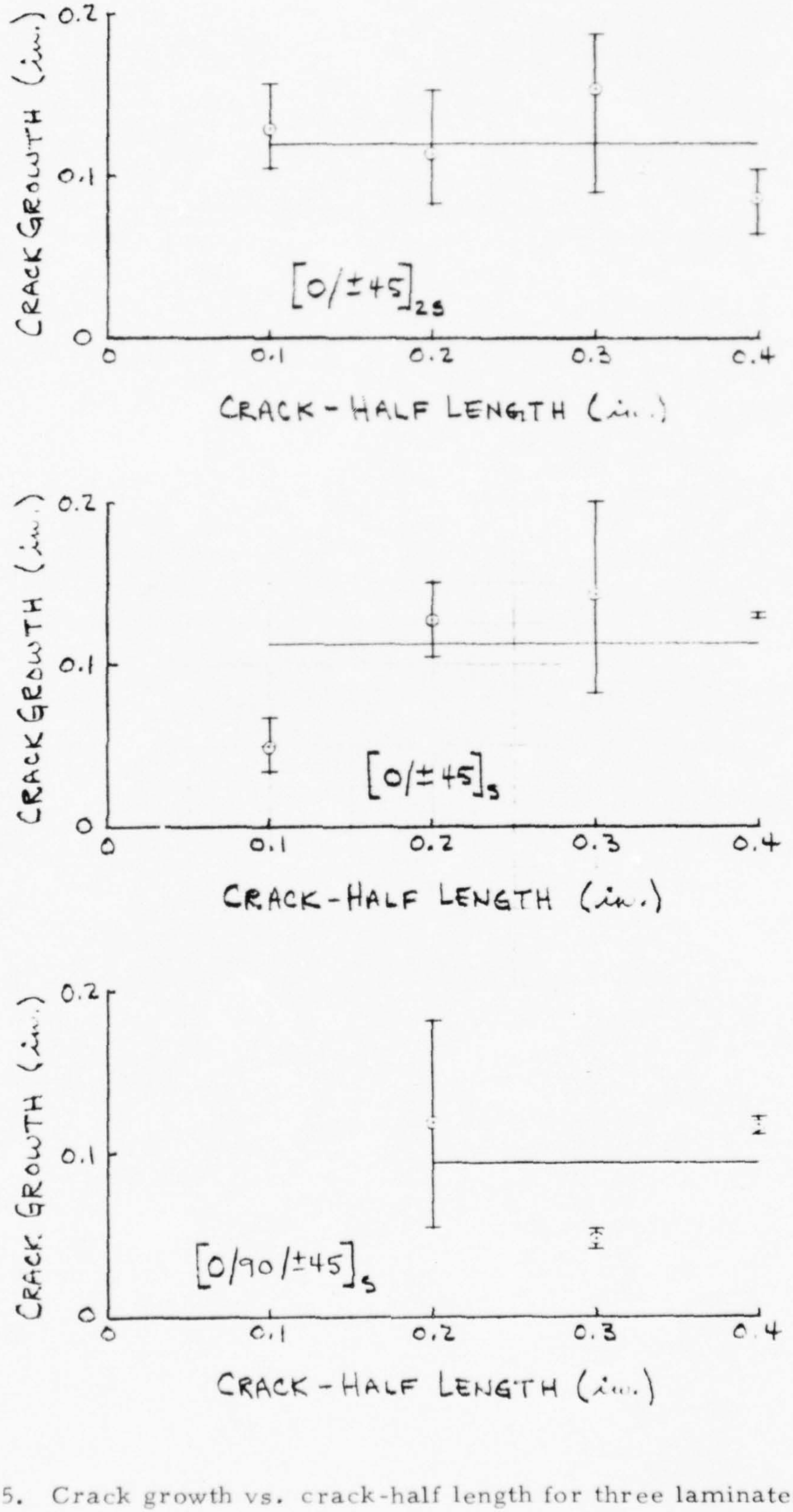


Fig. 15. Crack growth vs. crack-half length for three laminates

1975

ASEE - USAF SUMMER FACULTY RESEARCH PROGRAM

WRIGHT-PATTERSON AFB, OHIO

&

EGLIN AFB, FLORIDA

(CONDUCTED BY AUBURN UNIVERSITY)

OPTICAL PROPERTIES OF EUROPIUM-DOPED POTTASIUM CHLORIDE LASER WINDOW MATERIALS

Prepared By:	Thomas G. Stoebe, Ph.D.
Academic Rank:	Professor
Department and University:	Department of Mining, Metallurgical and Ceramic Engineering University of Washington
Assignment:	
(Laboratory)	Materials
(Division)	Electromagnetic Materials
(Branch)	Laser & Optical Materials
USAF Research Colleague:	G. Edward Kuhl
Date:	15 August 1975
Contract No.:	F44620-75-C-0031

OPTICAL PROPERTIES OF EUROPIUM-DOPED POTTASIUM CHLORIDE LASER WINDOW MATERIALS

by

Thomas G. Stoebe

ABSTRACT

Optical properties of samples from a large KCl:Eu crystal, produced as part of the AFML laser window development program, have been determined at wavelengths from 190 nm in the near ultraviolet to 20μm in the infrared region of the spectrum. The absorption bands due to Eu^{++}, located near 240 and 330 nm, are used to monitor the Eu-ion content, which is formed in general to behave as expected, but which shows anomalies from one side of the crystal to the other, perhaps due to a non-uniform temperature distribution during growth. Annealing is found to have little effect on Eu^{++} content, but irradiation to 10^7R causes a small decrease in the Eu^{++} absorption bands and the appearance of new bands, perhaps related to Eu-ions in another valence state.

In the infrared region, absorption bands not previously reported are observed in these KCl:Eu samples. These absorption bands are relatively weak in the as-grown samples, but are enhanced by irradiation. These absorption bands, which appear between 3.4 and 12.5μm, can explain the observed background absorption at the 10.6μm operating wavelength of the CO_2 laser. Further work is needed to identify and eliminate the causes of these absorptions.

I. INTRODUCTION

Alkali halide materials are currently under development for use as windows in high power CO_2 laser systems. Alkali halides have been chosen because they exhibit a high degree of transparency in the infrared region of the spectrum, with theoretical absorption coefficients below 10^{-4} cm^{-1} at 10.6μm[1], the operating wavelength of the CO_2 laser. Alkali halide can be produced, in relatively high purity or with controlled impurity dopants, as large single crystals. Since these single crystals are relatively weak, however, processes have been developed whereby a single crystal starting cylinder may be compressively forged into a high strength laser window.[2]

Of the alkali halide systems studied for CO_2 laser window use, pottasium chloride (KCl) has been found to have an optimum combination of transparency, forgeability and strength. The two systems now under advanced development are europium-doped KCl (KCl:Eu) and rubidium-doped KCl (KCl:Rb). These two systems have been shown to produce most consistantly good quality windows when reasonable care is taken in during the crystal growth and forging steps. Of these two systems, KCl:Eu has been chosen for further development work, while KCl:Rb is being held as a standby material.

While reasonable characterization of the various optical properties of KCl in general, and of KCl:Eu in particular, has been undertaken during the development process, few detailed investigations of the properties of this material have taken place. Ideally, the optical properties and the crystal lattice defects related to these properties, should be characterized in the KCl:Eu window materials at all stages in the production process to determine the influence of the Eu ions and of each production step (crystal growth and forging) on the properties of the final window material; an understanding of these properties and property changes could then lead to improvements in the production process for future generations of these windows.

Factors relating to the use and reliability of laser windows depend not only on optical transmission and strength, but also to the influence of operating conditions and environment on properties. Since even the theoretical power absorption of the window is not zero, some heating will take place in the window, causing defect annealing. The windows are protected from the gaseous environments by thin film coatings[3], but may need to be protected from damage caused by x-ray or gamma-ray irradiation.

This report deals with the beginning stages of an investigation of the properties of KCl:Eu laser window materials, as influenced by impurity state and distribution and by gamma-irradiation conditions. The specific scope and objectives of the project are discussed below.

II. OBJECTIVES AND SCOPE

This project involves the characterization of defect properties throughout one large KCl:Eu crystal produced for laser window development. Specifically, it was set out to determine

1. The optical absorptions due to Eu in this crystal throughout the ultraviolet, visible and infrared regions;

2. The variations in Eu content through the single crystal by optical means, to be compared with chemical means;

3. Any changes in the valence of Eu from its normal (original) state as Eu^{+2};

4. The influence of annealing on Eu content, valence and distribution; and

5. The effect of gamma-radiation on the absorption throughout the spectrum and on the Eu absorptions in particular.

Since KCl has been studied extensively in the literature[4], the basic properties and behavior expected are known, expecially in terms of optical defects in the visible region of the spectrum. This has aided in the direction of this short research program to those areas where this KCl:Eu window materials may differ from other KCl materials studied previously. The overall scope of the problem was aimed at determining those factors which may influence fabrication and operation of this specific material for laser window use at 10.6μm.

III. BACKGROUND

Alkali halides are a widely used class of materials for infrared optical instruments and related applications. They are cubic in crystal structure and are optically isotropic. They are easily grown from the melt as large single crystals; for demonstration purposes, the Harshaw Chemical Company has grown KCl single crystal boules as large as 32in. diameter. As the size of the single crystal increases, however, small angle boundaries are introduced which become macro-grain boundaries in crystals over 5 to 10 in. diameter.

A. Properties of KCl

KCl is a widely used and studied alkali halide material, which shows theoretical optical transparency from its electronic band edge in the vacuum ultraviolet to its infrared cut-off beyond 12μm. KCl is not hydroscopic, but reacts with water vapor to form a surface layer of hydroxide and related ions, and must therefore be handled carefully and stored in a dry place.

Pure KCl single crystals are quite weak, with a yield stress in bending or compression below 500 pounds/in^2(p.s.i.). Strengthening may be accomplished by alloying (doping with impurities), by irradiation, and by introducing grain boundarie The laser window developemnt projects[2,5,6], have used a combination of doping and grain boundaries(introduced by hot, compressive deformation or "forging") to improve the yield strength to over 4000psi. Radiation strengthening has been noted to increase the 10.6μm absorption and is susceptible to annealing above room temperature, and is probably not satisfactory for this purpose.

The properties of interest in a material which is to be used as a laser window are generally caused by impurities and other point defects, by line defects (dislocations) and by surfaces and grain boundaries. Of primary interest for the current studies are the point defects, including impurities, vacancies and vacancy aggregates, some of which are optically active defects which give rise to energy absorptions at specific wavelengths. Properties due to these defects, methods for their study and previous results in KCl are discussed briefly below.

B. Point defects of KCl

High purity single crystals of KCl will contain, due to statistical thermodynamical considerations,[8] an equilibrium concentration of Schottky defects, consisting of equal numbers of cation (pottasium) vacancies and anion (chlorine) vacancies. The concentration of these defects is given by

$$[V_K]\ [V_{Cl}] = A\exp -(H_s/kt) \qquad (1)$$

where V_K and V_{Cl} indicate vacancies on K and Cl sites, respectively, and [ ] represents concentration in mole fraction; H_s is the formation enthalpy for Schottky defects in KCl (best current value[9] is 2.52ev), k is Boltzmann's constant, T is temperature in °K, and A is a constant containing the formation entropy of the defect.

Europium is normally introduced into KCl in its divalent state as Eu^{++}. When KCl is intentionally doped with such a divalent cation impurity, cation vacancies are added to maintain charge neutrality, according to the charge balance equation,

$$2K^{+} \rightleftarrows Eu^{++} + V_K \qquad (2)$$

While the vacancy concentration at high temperatures will be controlled intrinsically by equation (1), at lower temperatures the intrinsic vacancy concentration will decrease below the extrinsic vacancy content defined by

$$[V_K] = [Eu^{++}] \qquad (3)$$

Hence the overall cation vacancy concentration will vary with temperature as shown in Figure 1. As the temperature decreases farther, the cation vacancies and divalent impurities are attracted to one another, forming impurity-vacancy dipoles. At still lower temperatures, these dipoles are attracted to one another, forming groups of two dipoles (called dimers) then after longer times, groups of three dipoles (trimers)[10].

Impurity-vacancy dipoles and their aggregation have been studied in KCl:Eu using ionic thermo currents[11] (ITC) and electron spin resonance 12-14 (ESR). Samples rapidly quenched to room temperature from 500°C, to assure a large metastable concentration of Eu++ - vacancy dipoles, showed a dipole content[11] of $1.8 \times 10^{18}/cm^3$. These dipoles react with one another and precipitate from solution in times varying from 500 hours at 40°C to 100 hours at 70°C. ESR results are consistant with this; the typical dipole spectrum is observed at and below 100°C, after quenching, in short time experiments.

Divalent cation impurities have been observed to change defect-related properties in most alkali halides. Optical absorption and luminescence, as well as thermo-luminescent behavior, are generally altered by the addition of extrinsic defects.[15] Of great importance in laser window development is the strengthening of the material caused by the addition of the divalent impurities. Such strengthening effects have been known for many years and studied by many investigators; Honeywell data[18] for several impurities in KCl is given in Fig. 2. Divalent impurities cause strengthening through the interaction of impurity-vacancy dipoles with dislocations, while monovalent impurities only interact through strain interactions caused by their difference in size as compared with K-ions; Hence, divalent ions cause strengthening much faster than monovalent ion impurities. Rb concentrations of over 1 percent are needed to produce the same strengthening as 100 parts per million (ppm) of Eu++ in KCl.

Europium ions exist generally either as divalent Eu++ ions, as discussed above, or as trivalent ions, Eu+++. The latter impurity would require an additional vacancy for charge compensation, would be expected to harden the crystal more than Eu++. It would induce different optical behavior in the KCl crystal than Eu++. Since trivalent impurities are little studied in alkali halides, their presence is eliminated by careful crystal preparation procedures if at all possible.

When divalent cation impurities are added to KCl, the total cation vacancy concentration increases, as noted above and in Figure 1. However, the statistical theory involved in equation (1) indicates that the product of the cation and anion vacancy concentrations must be constant,

$$[V_K] \cdot [V_{Cl}] = \text{const.} \qquad (4)$$

Thus when impurities enhance $[V_K]$, they correspondingly decrease $[V_{Cl}]$, and properties related to anion vacancies will decrease in the extrinsic region in such crystals.

C. Optical defects

All of the point defects discussed above, in their neutral state as well as when they act as electron or hole traps, have the potential of becoming optically active defects with optical absorption and luminescense bands at characteristic wavelengths throughout the electromagnetic spectrum. Since the goal of laser window research is to produce a material of minimum absorption at the laser wavelength (10.6μm in our case), it is essential that potential absorbing defects be eliminated.

Eu ions have characteristic absorption bands in the visible and ultraviolet region, and no added absorption has been noted at 10.6μm due to their presence. These absorption bands may be used to monitor the Eu ion content and valence state in KCl crystals.

The optical absorption spectrum of Eu++ as an impurity in KCl has been studied by several workers.[17-19] Two main absorption bands are observed at room temperature, centered around 243 nm and 326 nm in the near ultraviolet region.[17] These bands exhibit some structure that has been resolved at 77°K into peaks at 234, 240, 244, 251, 258, 329, 343 and 364 nm.[19] KCl:Eu++ has been shown to emit fluorescent blue color, as can be observed in large samples.

The absorption spectrum of trivalent Eu+++ ions has been observed only in europuim salts [20,21] and in $LaCl_3$:Eu[22]. The principal absorption lines in our region of interest are at 579, 526, 465 and 396 nm. Of these, the absorption at 465 has been reported as being about twice as intense as the others, while the presence of the absorption at 579 nm seems questionable. Kim and Moos [22] observed that Eu+++ was formed upon irradiation of $LaCl_3$:Eu++, and show that this process provides the electrons needed to form color centers in this crystal; optical bleaching of the color centers re-forms the Eu++ absorption bands in this material. Chowdari and Itoh[23] have observed a decrease in Eu++ absorption band intensity as a function or radiation time during x-ray exposure in KCl containing 10 ppm Eu++. This is interpreted as being caused by the conversion of Eu++ to Eu+++ by trapping a hole. Subsequent F-light bleaching partially restores the Eu++ absorption, through the trapping of an electron released during bleaching.

Unger and Perlman[11] observed the destruction of Eu++-vacancy dipoles by x-irradiation at room temperature. This was interpreted as being caused by (1) the capture of an electron by the Eu++, forming Eu+, or (2) the removal of the cation vacancy from the dipole by the conversion of the cation vacancy to an anion vacancy in an F-center formation process. Either of these mechanisms destroys the dipole, preventing dipole reorientation in an applied electric field (as in ITC).

The presence of impurities enhances the F-center formation rate, probably by the impurity (or the Eu - vacancy dipoles) trapping an interstitial Cl so the dipole would change its reorientation frequency, decreasing the apparent concentration, as observed by Unger and Perlman.[11] This should also cause the appearance of an ITC peak at a higher temperature (lower frequency). The latter has not yet been reported in KCl:Eu. Possibly related to the above is the observation by Parfianovich et.al.[24], that Eu++ vacancy dipoles are the predominant center for luminescence emission in KCl:Eu as well as for other Eu-doped alkali halides. Further investigations of the role of Eu++ vacancy dipoles in determining optical properties of KCl:Eu seem to be needed.

Deutsch[25], in studying possible impurity contributions to the 10.6μm absorption in KCl, found a weak absorption band near 9.7μm, which he concludes is due to an associated complex of binding energy near 0.1eV. By comparison, Deutsch suggests a complex analagous to CrO_4 = -anion vacancy pairs, which have an association energy of 0.3eV. Unfortunately, Deutsch does not determine the purity of his KCl crystals, which are identified only as "commercial"; the introduction of a major impurity such as Eu during crystal growth could drastically change the content of minor impurities in the resultant material. As a result, it cannot be determined if this absorption is relevant in KCl:Eu.

Recently, Jain et.al.[26] have compiled a compendium of experimentally observed atomic and molecular impurity absorptions in the alkali halides. Much of the same data has been evaluated by Sparks and Duthler[27] in terms of possible contributions of various molecular ions to the absorption at 10.6μm. The latter data is reproduced in Table I.

D. Color Centers

Under the influence of irradiation from ultraviolet light, x-rays, γ rays and nuclear particles, alkali halides become damaged through the production of color centers. Color centers are ionized point defects of various types, some of which are catalogued in Table II. The basic type of defect created by irradiation in KCl is the F-center, which has a strong absorption band centered near 540 nm in the visible region of the spectrum. The F-center is created during irradiation by the formation of an anion vacancy-interstitial anion pair. The energy required to remove the interstitial anion from the vicinity of the vacancy it creates is thought to be supplied by the electron-hole pair recombination energy.[4] In this process, the interstitial anion captures a hole and becomes an interstitial Cl° ion or H center, which exists as a split-interstitial (the interstitial Cl° and a substitutional Cl^- on one shared Cl - ion site). The vacancy left behind by the departure of the interstitial may then capture a hole made available by the irradiation, and become an F-center.

Table I. Experimentally observed absorption frequencies of several impurity ions in KCl crystals and the estimated impurity concentration to produce $=10^{-4}cm^{-1}$ at the CO_2 laser frequency (from Ref. 27)

Ion	Frequencies(cm^{-1})	Conc.(ppm) for $\beta(943cm^{-1})=10^{-4}cm^{-1}$
H^-	500	> 10
OH^-	3640	> 100
SH^-	2590	> 100
CN^-	2089	100
BO_2^-	590,1972	10-100
CO_2^-	1696	10-100
N_3^-	642,2049	100
NCO^-	629,1232,2169	10
NO_2^-	805,1290,1329	0.1
NO_3^-	842,1062,1396	10
CO_3^{2-} [a]	680-720,883,886,1058,1064,1378,1416,1488,1518	10
BO_3^{3-}	736,949(w),1222	1-10
SeO_3^{3-}	broad738-850	1-10
HCO_3^-	589,672,713,840,971,1218,1346,1701,3339	0.01
BH_4^-	1143,2321	10
SO_4^{2-} [a]	630,978,1083,1149,1188	0.03
CrO_4^{2-} [a]	862,890,930,941	< 0.1
SeO_4^{2-} [a]	834,860,909,923	< 0.1
MnO_4^{2-} [a]	899,909,914,925	< 0.1
NH_4^{2+}	1405,3100	10

a. Exact frequencies and intensities very dependent on presence of compensating M^{2+}.

Table II. Color Centers in Alkali Halides

Description of Defect	Notations used
Anion vacancy (uncharged)	F^+, α
Anion vacancy with one electron	F
Anion vacancy with two electrons	F^-, F'
Two adjacent F centers	F_2, M
Three adjacent F centers	F_3, R
Ionized M center	F_2^+, M^+
F center adjacent to monovalent cation impurity	F_A
F center adjacent to divalent cation impurity	F_z, Z
Intersitial halogen atom	
(exists as a molecular ion, two anions at one anion site)	H, Cl°
Self-trapped hole (one anion at one anion site)	V_k

During room temperature x-irradiation of relatively pure KCl samples, it is found that an equilibrium is established between the F-center concentration and the concentration of those centers composed of aggregates of F-centers.[28] The model of the M-center (two nearest neighbor F-centers) is well established and the concentration of R-centers is proportional to the product of the concentrations of F-& M-centers, indicating its identity as three nearest neighbor F-centers. Additional centers exist with intensities related in the same manner: the concentration of N_2-centers, showing an absorption near 1μm is proportional to the product of the concentrations of the F-& R-centers, indicating that the N_2 band is made up of four F-centers. At higher concentrations these relationships deviate from linearity indicating the presence of higher aggregates with absorptions at still longer wavelengths.

F-centers may be destroyed by illuminating the crystal with light at the same wavelength as the F-absorption band, a process called bleaching.[29] This transforms F-centers into F-aggregate centers and some of the other centers which are noted in Table I. Many of these centers are not stable at room temperature and the effect of irradiation is thus annealed out. Other centers, such as M, R and Z-centers, anneal only at elevated temperatures.

The intensity of the 10.6μm absorption band in KCl has been related indirectly to color center content by Lipson[30], et.al. Using high exposures to gamma radiation (10^7-10^8 R), these authors found an increase in the 10.6μm absorption, measured using laser calorimetry, roughly proportional to the concentration of M-centers in the crystals. The increased absorption is postulated to be caused by the long wings exhibited by the M band[7] and possibly by others of the F-aggregate centers. Work by Magee, et.al.[30] has indeed shown a further correlation between the increase in 10.6μm absorption caused by irradiation and the concentration of R-centers in KCl:Eu crystals. Using additive coloration, which increases F- and F-aggregate center concentrations without an increase in Cl interstitials (H centers), Lipson, et.al.[7] further demonstrate that the increased absorption is not related to H-center concentration.

IV. METHODS

The experimental methods used in this project to determine optical properties were principally measurements of absorbance or transmittance of polished samples, carried out at room temperature. Instruments used were a Cary 14 spectrometer, used for wavelengths from 190 nm in the ultraviolet to 3μm in the near infrared, and Perkin-Elmer spectrometers, models 625 and 525, between 3μm and the infrared cut off (near 20μm).

The absorbance value of some of the absorption bands observed in these samples was beyond the absorbance value of 2 which is the maximum absorbance measureable with the Cary 14. To allow measurements to be performed at higher absorbance, a set of neutral density filters were used, calibrated over the wavelength ranges of interest. In the normal operating mode, this causes an abnormal opening of the slits and a loss of resolution at the peaks of the absorption bands, an effect that can be corrected by increasing the lamp intensity and increasing the slit width control. Most of the absorption curves were obtained in the normal operating mode, but when greater resolution was desired, greater control over the slit openings was exercised; when the slit control was used, it is noted in the results shown.

The transmittance T, reflectance R, and absorbance A of a sample are related by [32]

$$T = \frac{(1-R)^2 \; e^{-\beta x}}{1-R^2 \; e^{-2\beta x}} \qquad (5)$$

$$\log_{10} \frac{1}{T} = A = \log_{10} \frac{e^{\beta x}}{(1-R)^2} \qquad (6)$$

where β is the absorption coefficient, and x is the distance into the sample. Accordingly,

$$A = \log_{10} \frac{1}{(1-R)^2} + (\beta x) \; \log_{10} e \qquad (7)$$

where $R = (n-1/n+1)^2$, n being the index of refraction of the material. Since the first term in equation (7) is simply the background absorbance of the material, the peak height above background is related to the second term, allowing β to be calculated.

Once the absorption coefficient for an absorption band is known, the concentration of centers causing that absorption may be calculated using Smakula's equation:[33] The concentration in number per cm^3, N, is related to β in cm^{-1} at the maximum absorbance, β_{max}, by

$$Nf = 1.29 \times 10^{17} \; \frac{n}{(n^2+2)^2} \; \beta_{max} \; W \qquad (8)$$

where f is the oscillator strength (related to the probability of the optical transition causing the absorption), n is the index of refraction, and W is the half width of the absorption in electron volts. If the oscillator strength is known or can be determined by calibration, the concentration of absorbing centers may be calculated.

Samples for the current set of experiments were obtained from a Harshaw-grown 17" diameter, 6" high KCl:Eu++ crystal containing numerous large-angle grain boundaries. The crystal showed a fluorescent blue color due to the Eu impurities contained therein. This crystal, which had been delivered to Honeywell as part of the laser window development project[2], was denoted KC.01ECH97. The crystal was sectioned by Honeywell according to the plan shown in Fig. 3. Samples were then extracted from the upper and lower slices, as shown in Fig. 4, and labeled according to the plan in Fig. 5; samples H1-H4 were further displaced 4cm to the left due to a fracture. Samples from the first end of the crystal to solidify (the bottom in Fig. 3) are referred to as "cone" samples and samples from the opposite end are called "heel" samples. The small samples used were all 1cm square by 0.5cm thick. Samples were finished by water polishing the edges and mechanically polishing the faces to an optical finish using 0.3μm alumina polishing powder.

To differentiate between properties related to the Eu impurities and those related to the host KCl lattice, further samples of commercial purity KCl optical windows were tested. These samples were standard Harshaw optical windows of 0.5 cm thickness, and are designated 1A and 2A in this report.

Radiation experiments were performed by exposing the required samples to Co^{60} gamma rays using an Isotopes Specialties Co. radiation source with an exposure rate of 6.9×10^4 R/hr. Radiations were performed in air at room temperature for times between 1.45 hr. (10^5 R exposure) and 153 hr. (1.1×10^7 R exposure).

Annealing experiments were performed by heating samples in air at 650°C for times between two and six hours. Samples were either cooled in the furnace by cutting off the furnace power (furnace cooled) or cooled rapidly by placing the sample on an aluminum block (air cooled). The surfaces of these samples clouded by reaction with the atmosphere and were subsequently re-polished using 1μm and 0.3μm alumina powder.

Samples were stored in a desicator and were handled by tweezers or by gloved hands. All experiments were performed in air with time outside the desicator kept as short as possible.

V. RESULTS AND DISCUSSION

A. Optical Absorption in KCl:Eu

1. Eu^{++} Absorption in As-received Crystals

The absorption curves due to Eu^{++} ions in KCl:Eu samples taken from six locations throughout the crystal were determined at room temperature using the Cary 14 spectrometer in its ultraviolet and visible modes. Two composite absorption curves are shown in Fig. 6. These absorption curves were determined using minimum slit widths for maximum resolution of the absorption peaks. The absorption curve for sample C4, taken from the cone section of the sample (see locations in Fig. 7), shows less total intensity than does that of sample H 11, taken from the heel, due to the increased Eu^{++} content in the heel of the sample, discussed below.

The wavelengths of the absorption bands in KCl:Eu, reported in the literature for experiments at 77°K where individual peaks are better resolved, are indicated across the bottom of Fig. 6 at 234, 240, 244, 251, 258, 329, 343 and 364 nm. The single peak near 240 nm is the result of the sum of the five bands indicated, which are unresolved at room temperature. The three bands at 329, 343 and 364 nm are partially resolved, the latter two appearing as shoulders on the more intense peak observed near 330 nm.

The absorption coefficients for the peaks at 240 and 330 nm are reported in Fig. 7 for the six sample locations indicated. The absorption coefficients were determined using equation (7), with the sample background observed at 500 nm (beyond the wings of the curves in Fig. 6) taken as the first term in the equation (which is subtracted), and adjusting the peak heights to the observed instrumental baseline shown in Fig. 6. At three of the locations, the observed absorption coefficient values are averages of three or four samples, as indicated (H1-4, C1-4, H10-12). In these cases, the observed variation is within ±3 percent at each location, indicating that the observed variations in absorption coefficient are real.

The observed variations in absorption coefficient may be discussed in terms of the distribution of Eu^{++} during the crystal growth process. During initial solidification (from the cone end), Eu^{++} is rejected from the solid and remains in the liquid. As the liquid continues to solidify, its higher Eu^{++} content causes increased Eu^{++} content in the solid as solidification proceeds. The last parts of the sample to solidify are quite high in Eu^{++} content and are discarded. In the current sample, since the sides of the mold cool first, the parts of the crystal at samples C1-4 and at C12 will solidify first, followed by that at sample C8; near the end of the solidification process, at a higher Eu^{++} content, samples at H1-4, then H10-12, then H5 will solidify. From this we expect to see a decrease in Eu^{++} content from the center to the side of the crystal, as is noted, and from bottom to top, as is also noted.

While the observed general concentration gradients are expected, two anomalies exist in the specific data: (1) sample C12 shows a lower content of Eu^{++} than does C1-4, even though these sample locations are arranged symmetically about the sample center line, and (2) samples H10-12 show a lower Eu^{++} content than do H1-4, even though samples H1-4 are nearer the side (and hence should show a lower content). These results could be caused by a non-uniform temperature distribution during crystal growth, or could be due to large concentration fluctuations.

Honeywell chemical analysis data for a similar (but smaller) crystal is given in Fig. 8. This data shows the same general behavior explained earlier, and shows considerable fluctuations in concentration across the diameter of the sample.

Honeywell has concluded that the observed composition fluctuations are due to Eu^{++} segregating at grain boundaries, and presents chemical analysis data of regions of the crystal known to contain grain boundaries to confirm this hypothesis.

In the current absorption experiments, samples containing grain boundaries (i.e. sample C1, see Fig. 5) were tested to see if higher concentrations of Eu^{++} could be observed. The results were negative, in that the absorbance values from sample C1 are identical to that from samples C2-4 to within ±3 percent. Since none of the other samples tested in the present work contained grain boundaries (see Fig. 5), we must conclude that the anomalous concentration distribution observed in our crystal is not caused segregation to grain boundaries. Indeed, the data presented in Fig. 7 indicates an overall concentration increase on the left side of the crystal, which could well be explained by a non-uniform temperature gradient during crystal growth.

Additional characterization of as-received samples is available from mechanical properties tests and infrared absorption at 10.6μm. Values of yield stress σ and absorption coefficient β at 10.6μm are shown for reference in Fig. 8. For our present crystal, values obtained by Honeywell are σ = 1205 psi (heel) and 945 psi (cone); β at 10.6μm has not been determined.

2. Effects of Annealing on Optical Absorption

The Honeywell development work has indicated that Eu-doped KCl becomes more brittle after anneal in air at 650°C, whereas other doped KCl samples became more ductile after annealing.[2] Honeywell speculated that this was caused by a valence change Eu^{++} to Eu^{+++} caused by annealing KCl: Eu in an oxidizing atmosphere.

We have observed the effect of annealing on the major Eu^{++} absorption peaks. Samples H1 and H2 were annealed (6 hr.) and one was air cooled while the other was furnace cooled. A modest decrease in absorption coefficient (4 percent) was observed in the 240 nm peak for both samples, while the 330 nm peak remained unchanged. Similar results were obtained upon annealing another sample for 12 hours. No changes in overall absorption band shape was seen, nor were any other new absorptions visible. Hence, we can conclude that any change in Eu^{++} content is small and virtually undetected in these experiments.

3. Infrared Absorption in KCl:Eu

The principal interest in absorptions in laser window materials lies in peaks which may appear near the operating frequency. The absorptions discussed above are used here for a characterization of the Eu^{++} ion in the KCl crystals and provide needed background information for studying infrared absorptions that may be Eu-related.

Infrared absorption was measured over a wide range of frequencies in many of our samples, and several absorption peaks were noted. These peaks are tabulated in Table III. All were low in intensity and could be seen only with a 10X scale expansion. Care was taken to verify the presence of the peaks, since the low intensities and instrumental peculiarities could cause difficulties in the differentiation of the absorption peaks from the backgrounds. Fortunately, it was found that irradiation enhances the observed peaks, producing more definitive evidence of their existance. This data is discussed in more detail in section B3 below.

Table III. Infrared absorptions seen in KCl

Sample	Peak Frequency, cm^{-1} (with relative intensities)					
	798	1014	1255	1728	2850	2918
Pure (2A), unirradiated	-	-	-	-	?	?
(1A), irradiated 10^7R	-	-	-	-	2(?)	2(?)
Doped (H12), unirradiated	-	2(?)	3(?)	4	4	6
(H10), irradiated 10^7R	5	5	8	6	8	10

B. Optical Absorption in Irradiated KCl

1. Eu^{++} Absorptions in Irradiated Crystals

The influence of γ-irradiation on the Eu^{++} absorption bands in sample H10 is shown in Fig. 9. Here the absorption curves are shown for 190-300 nm using the ultraviolet mode of the Cary 14 in Fig. 9A, and for 300-420 nm using the visible mode in Fig. 9B. The general features caused by the irradiation are a decrease in peak height, a general increase in background level, and the growth of at least one additional peak, near 190 nm. The changes in absorption coefficient due to irradiation and subsequent optical bleaching (shown below) are given in Table IV, where the effect of irradiation is seen to be a 7 to 8 percent decrease in peak absorption coefficient. Bleaching seems to make no significant changes in the absorption coefficient.

These results are generally similar to those of Chowdari, et.al., who report a decrease in the 330 nm absorption with x-irradiation, and a (partial) restoration of this decrease during F-band bleaching.[19] Chowdari, et.al., interpret these changes as due to the conversion of a small portion of Eu^{++} to Eu^{+++} during irradiation and the reverse during F-band bleaching.

Table IV. Absorption coefficient at 240 nm and 330 nm (Eu^{++} bands) after irradiation and optical bleaching in sample H10

Step	β, cm^{-1}	
	240 nm	330 nm
As-received	12.21	11.16
Irradiated to 10^5R	11.98	10.97
Irradiated to 10^6R	11.61	10.51
Irradiated to 10^7R	11.29	10.32
Bleached 30 min.	11.11	10.37
Bleached 75 min.	11.01	10.28
Bleached >2 hrs.	10.92	10.05

2. Color Centers in Irradiated Crystals

Irradiation of our KCl:Eu samples produces an F-absorption band near 560 nm and smaller F-aggregate bands at longer wavelengths. The F-band is quite intense (absorbance >2) even after 10^4R, but the F-aggregate centers become well developed only at 10^7R. The development of the F-aggregate centers is shown in Fig. 10, where a well developed M band is seen after 10^7R near 830 nm; R-bands are seen near 740 nm and N-bands are indicated near 1000 nm.

The F-absorption band shape after 10^7R is shown in Fig. 11. This band has an obviously complex shape, which was not observed at 10^5R and was only partially developed after 10^6R.

During the determination of infrared absorption in the samples after irradiation, bleaching of the F and F aggregate bands was observed due to illumination by the glow-bar source in the Perkin-Elmer spectrometer. After bleaching for 30 minutes by this broad band light source, the F-band bleached to a more normal shape, as shown in Fig. 10. By taking the difference between these two curves it is seen that the abnormal F-band shape is caused by two additional bands that form on irradiation. These new bands show maxima near 520 nm and 590 nm, although these peak positions could be in error by as much as 20 nm due to this method of determination.

The F-absorption band in pure KCl crystals has a shape quite similar to that shown after bleaching in Fig. 10; the comparison of the F band in the pure sample with that in the doped sample also indicates the presence of new bands near 520 and 590 nm.

While the identity of these new bands formed on irradiation cannot be determined exactly, the peak positions are near two of the positions indicated in the literature as Eu^{+++} absorption wavelengths, namely 526 and 578 nm. However, the third Eu^{+++} absorption expected in this region, at 426 nm, is not observed. Another possibility is absorption due to Eu^{+}, discussed in NaCl by Gorobets, et.al.[34] That these bands are related in some way to Eu-ions is indicated by the presence of these bands only in Eu-doped samples.

An additional band formed by irradiation appears near 190 nm (Fig. 9). While Chowdari, et.al.[19] observe the formation of a band in KCl:Eu at 223 nm, this latter band is not seen in our samples, and it is doubtful if this is the same as the band we observe at 190 nm.

3. Infrared Absorption in Irradiated Samples

The infrared absorption spectrum for sample H10, irradiated to 10^7R, is shown under normal instrument conditions in Fig. 12A. Little significant absorption is observed, although hints of small absorption bands may be noted. The absorption spectrum of this same sample is shown at 10x scale expansion in Fig. 12B, curve a, where 6 absorption bands are observed. This curve is compared with similar sample H12, unirradiated, taken at 10x scale expansion shown in curve c, and the observed absorptions are tabulated in Table III.

It may be noted in Fig. 12B that each of the small absorption bands noted in the unirradiated sample is present and shows an enhanced intensity in the irradiated sample, and that one additional band is observed at $798cm^{-1}$. The bands at 1014, 1255 and 1728 cm^{-1} are quite weak in the unirridiated sample, (curve c); their presence was confirmed only by comparison with the irradiated sample (curve a). The only absorptions noted in the undoped samples 1A and 2A were those at 2850 and 2918 cm^{-1}; the presence of these peaks is questionable in the unirradiated sample, but is obvious (although small) in the irradiated sample (see Table III).

All of these infrared absorptions must be caused by impurity ions, molecules or complexes, but their identities are not known at present. The peaks at 2850 and 2918, since they are present to some extent in all samples tested, could be caused by a trace impurity present in all of Harshaw's KCl starting material. The other peaks, observed only in Eu-doped samples, may be related to Eu or to another impurity introduced during doping or growth of this KCl:Eu crystal. The cause of the radiation-enhancement of the observed bands is also not clear at present. Radiation-induced absorptions were expected due to previous results by Phillipi[35] on KCl irradiated by and supplied by the SRI group[30]. However, only one of the observed bands (1255 cm^{-1}) matches Phillipi's observed absorption bands and the intensities we observe are much less than his. These differences are perhaps not suprising since a different batch of Harshaw pure KCl was used in the SRI work.

The presence of infrared absorption bands in the KCl:Eu window material studied here is of extreme importance, since the tails of the absorptions observed near 10.6μm could cause directly the observed 10.6μm absorption in KCl:Eu. It is clearly necessary to identify and eliminate these absorptions in the laser window material.

That radiation enhances these absorption bands is also a significant observation. First, it emphasizes the need to eliminate these absorption bands if the laser window is ever to operate near x-or y-radiation fields. Second, it demonstrates that prior explanations for radiation enhancement of 10.6μm absorption, which were related to the extrapolation of the M-or R-center wings,[25,30] are not applicable in this material; rather, it may be related directly to the growth of the bands observed near 10 and 12.5μm. Finally, this radiation enhancement provides an additional method for the study of the weak bands in the unirradiated material.

VI. CONCLUSIONS AND RECOMMENDATIONS

The most important result of this research project is the observation that infrared absorption bands are present in this KCl:Eu laser window crystal. Such bands have not been reported previously.

The study of absorption bands in KCl:Eu was extremely helpful in establishing that these techniques may be used to monitor Eu^{++} content in these laser window materials. While annealing showed little effect, irradiation showed a small but significant decrease in Eu^{++} content and the growth of new bands, which may be related to Eu-ions in another valence state.

The results regarding the infrared absorptions presented in this report must be regarded as preliminary, since no detailed determinations of absorption coefficient could be made from the low intensity absorptions. Further work is essential to characterize and identify these absorptions, with the goal of eliminating their causes. Two methods for obtaining enhanced absorption are to use low temperature measurements (reducing the background absorptions) and to irradiate the samples to a higher dose, since the absorptions may be enhanced by irradiation. These experiments are planned at AFML in the near future.

Of additional interest is whether these absorptions are modified in the compressive forging operation, which is the next step in laser window production. It would also be useful to determine if these absorptions are present in KCl:Rb, the alternate window material; this determination could determine whether Eu-doping itself or another step in crystal growth causes introduction of these absorptions. It is recommended that research on these questions be undertaken.

REFERENCES

1. M. Sparks and L.J. Sham, Phys. Rev. B8, 3037 (1973).

2. W.B. Harrison, et.al., "Halide Material Processing for High-Power, Infrared Laser Windows", AFML Report, Honeywell, 1975.

3. "Protective-Antireflective Thin Film for Polycrystalline Zinc Selenide and Alkali Halide Laser Windows", AFML Report TR-75-18, Hughes Research Lab., Feb. 1975.

4. For discussions and references see the review by E. Sonder and W.A. Sibley, "Defect Creation by Radiation in Polar Crystals", in Point Defects in Solids, Vol 1, ed. J.H. Crawford and L.M. Slifkin, Plenum Press, New York, 1972, p. 201.

5. S.A. Kulin, et.al., "Development of Polycrystalline Alkali Halides by Strain Recrystallization for Use As High Energy Infrared Laser Windows", AFML Report TR-74-17, Manlabs, Jan 1974.

6. "Chemically Strengthened Polycrystalline Pottasium Chloride for High Power IR Laser Windows", AFML Report TR-74-165, Harshaw Chemical Co, Aug 1974.

7. H.G. Lipson, et.al., in Proc. Fourth Conference on High Power Laser Window Materials, Advanced Research Projects Agency, Jan 1975, p. 590.

8. See, for example, A. D. Franklin, "Statistical Thermodynamics of Point Defects in Crystals", in Point Defects in Solids, Vol. 1, op cit, p.1.

9. R.G. Fuller, "Ionic Conductivity", ibid, p. 103.

10. A discussion of dipole association with references is given in T.G. Stoebe and S. Watanabe, Phys. Stat. Sol. (a), May 1975.

11. S. Unger and M.M. Perlman, Phys. Rev. B6, 3973 (1972).

12. R. Rohrig, Phys. Lett. 16, 20 (1965).

13. P.G. Nair, et.al., J. Phys. Chem. Solids 29, 2183 (1968).

14. G. Aguilar, et.al., J. Chem. Phys. 60, 4665 (1974).

15. F.A. Kroger, Chemistry of Imperfect Crystals, North-Holland, Amsterdam, 1964.

16. "Halide Materia l Processing for High Power Infrared Laser Windows, Interim Technical Report No. 2, Honeywell, Feb. 1973.

17. A.K. Mehra, J. Opt. Soc. Amer. 58, 853 (1968).

18. R. Reisfeld and A. Glasner, J. Opt. Soc. Amer. 54, 331 (1964).

19. B.V.R. Chowdari and N. Itoh, Phys. Stat. Sol. (b) 46, 549 (1971).

20. K.H. Hellwege, et.al., Z. Phys. 148, 112 (1957).

21. L.G. DeShazer and G.H. Dieke, J. Chem. Phys. 38, 2190 (1963).

22. B.F. Kim and H.W. Moos, Phys. Rev. 161, 869 (1967).

23. B.V.R. Chowdari and N. Itoh, Phys. Stat. Sol. (b) 46, 549 (1971).

24. I.A. Parfianovich, E.I. Shuralever and P.S. Ivaknenko, J. Luminescence 1, 2, 657 (1970).

25. T.F. Deutsch, Appl. Phys. Lett. 25, 109 (1974).

26. S.C. Jain, et.al., "Electronic Absorption and Internal and External Vibrational Data of Atomic and Molecular Ions Doped in Alkali Halide Crystals", National Bureau of Standards Report NSRDS-NBS 52, July 1974.

27. M. Sparks and C. Duthler, Second Technical Report, "Theoretical Studies of High Power Infrared Laser Window Materials", Xonics, Dec. 1973.

28. S. Schnatterly and W.D. Compton, Phys. Rev. 135, A 227 (1964).

29. I. Schneider and H. Rabin, Phys. Rev. 140, A1983 (1965).

30. T.J. Magee, et.al., Phys. Stat. Sol. (a) 29, June 1975.

31. R.W. Warren, Phys. Rev. 155, 943 (1967).

32. H.J. Hrostowski, "Infrared Absorption in Semiconductors", in Semiconductors, ed. N.B. Hannay, Reinhold, New York, 1960, p. 437.

33. J.H. Schulman and W.D. Compton, Color Centers in Solids, Macmillan, New York, 1962.

34. B.S. Gorobets, et.al., Sov. Phys. - Doklady 13, 519 (1968).

35. C. Phillipi, AFML, private communication.

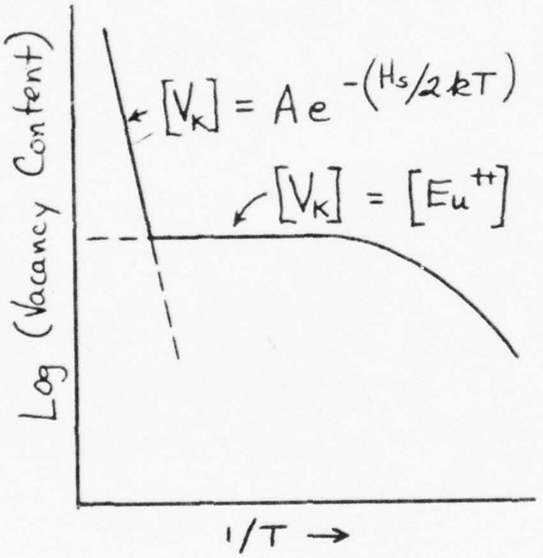


FIGURE 1. Temperature dependence of Vacancy concentration.

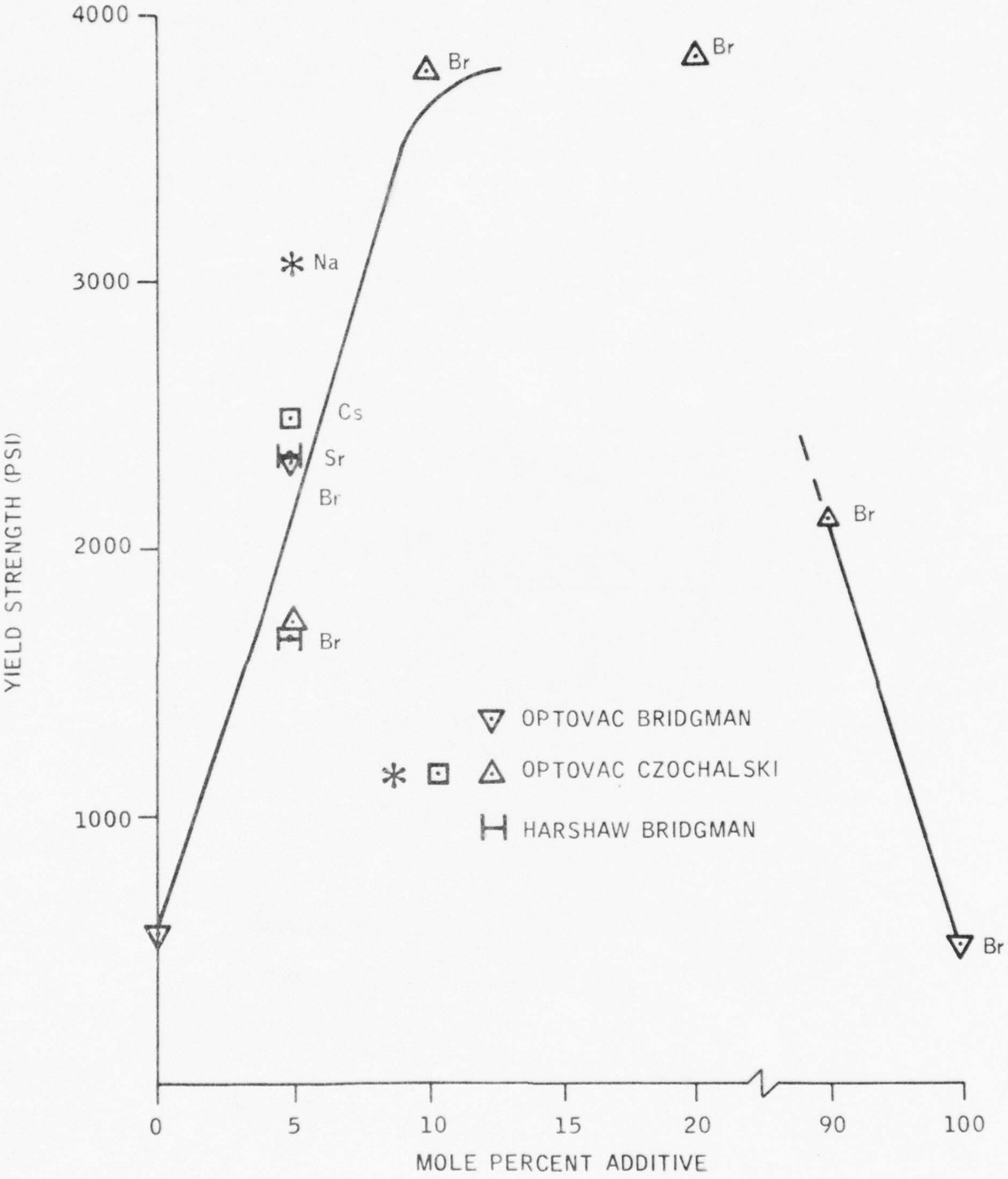


Figure 2. Influence of Na, Cs, Sr and Br Alloying Additions on KCl
(From Rof. 16)

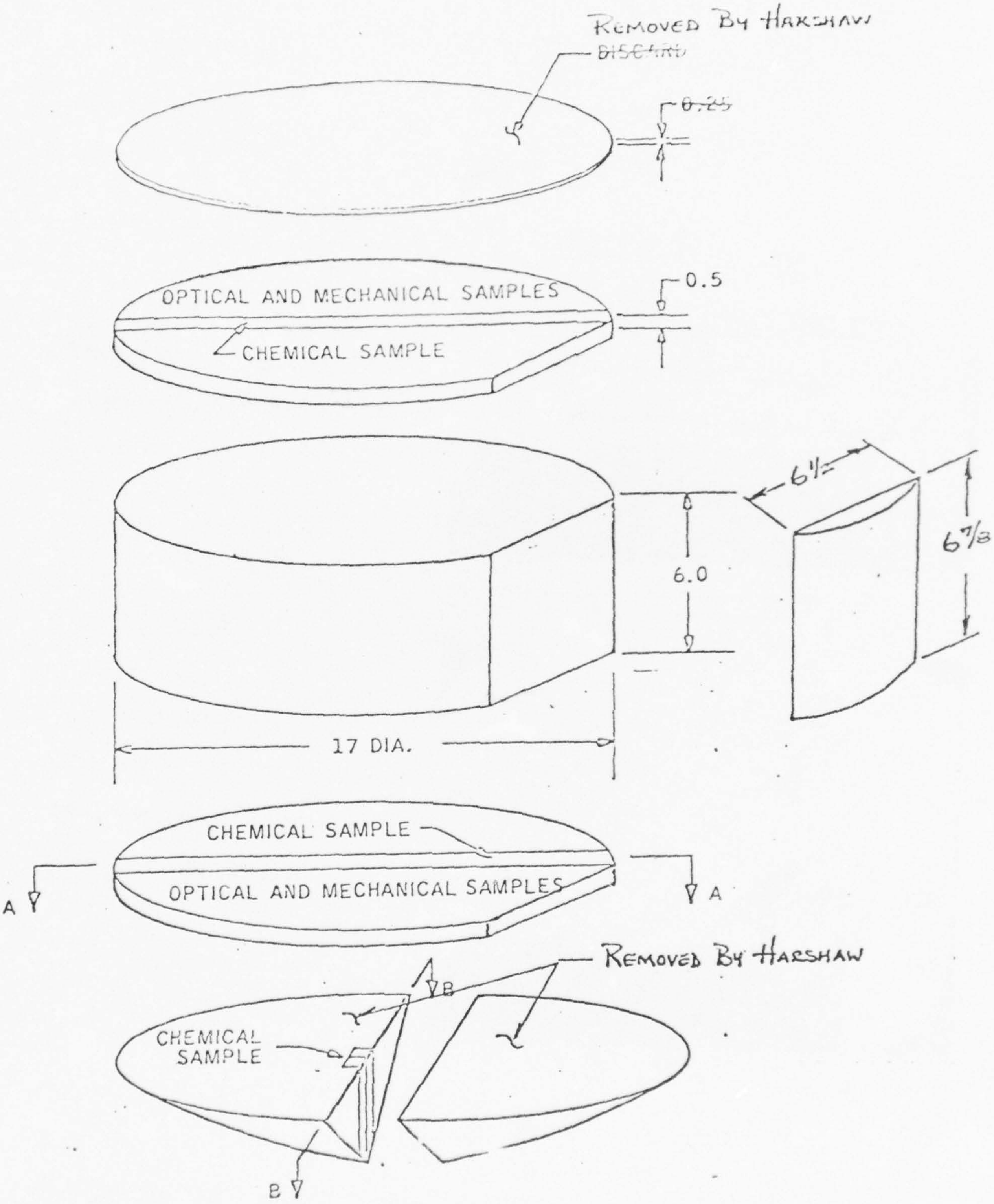
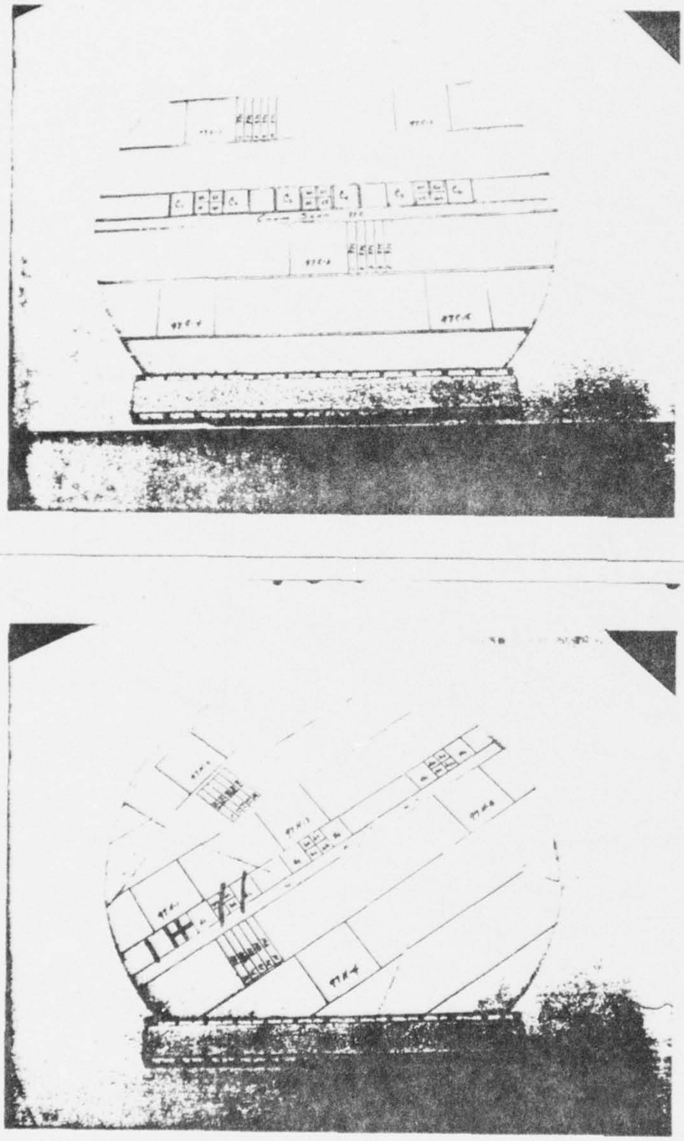


FIGURE 3. SECTIONING PLAN FOR KC.01ECH97

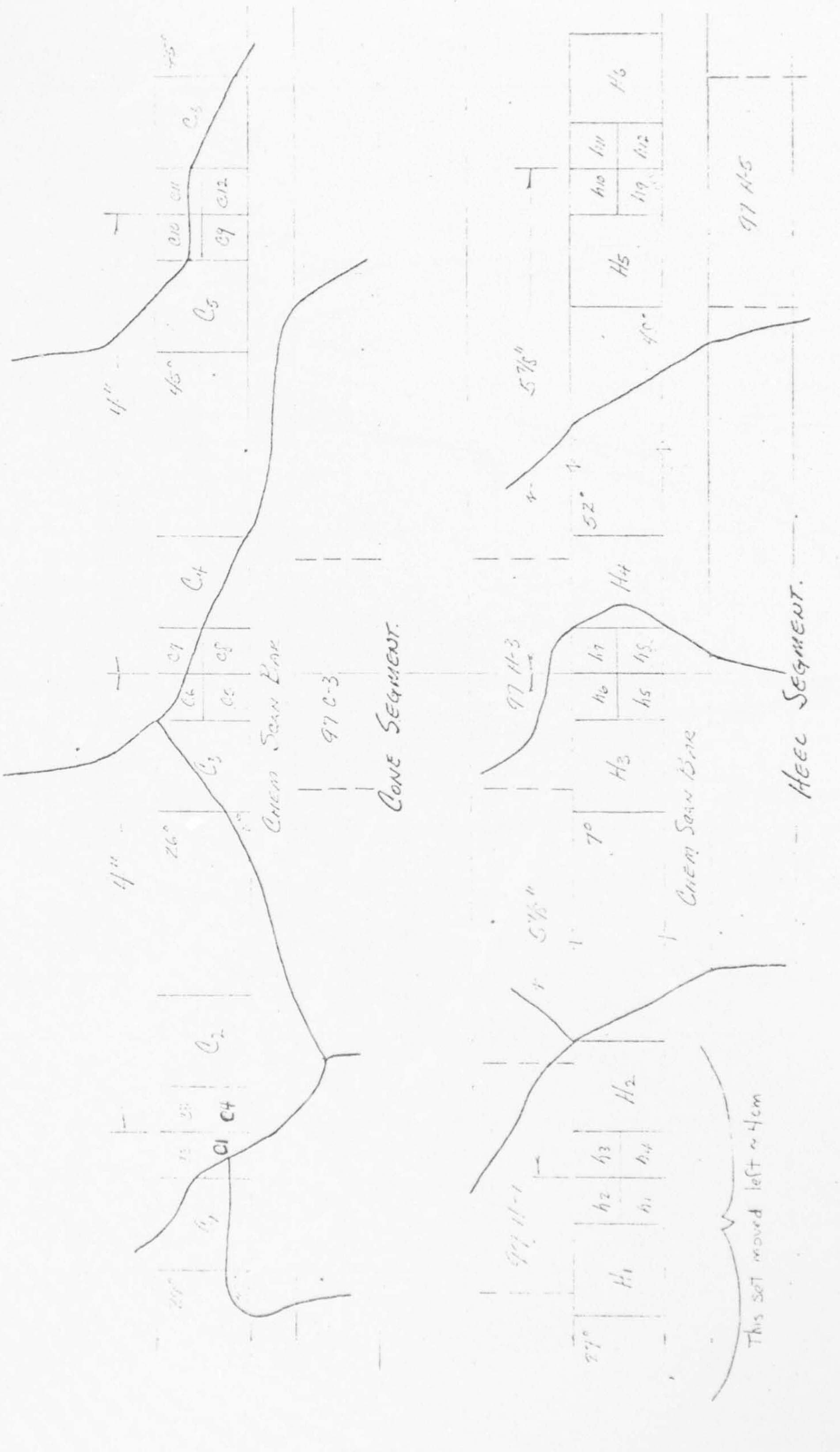


CONE SEGMENTS
(LOWER SLICE)

HEEL SEGMENTS
(UPPER SLICE)

FIGURE 4 : Sample Segments From KC.01 ECH 97

FIGURE 5 PLAN SHOWING MACROGRAIN BOUNDARIES AND CRYSTALLOGRAPHIC ORIENTATIONS RELATIVE TO CONE AND HEEL SAMPLES.



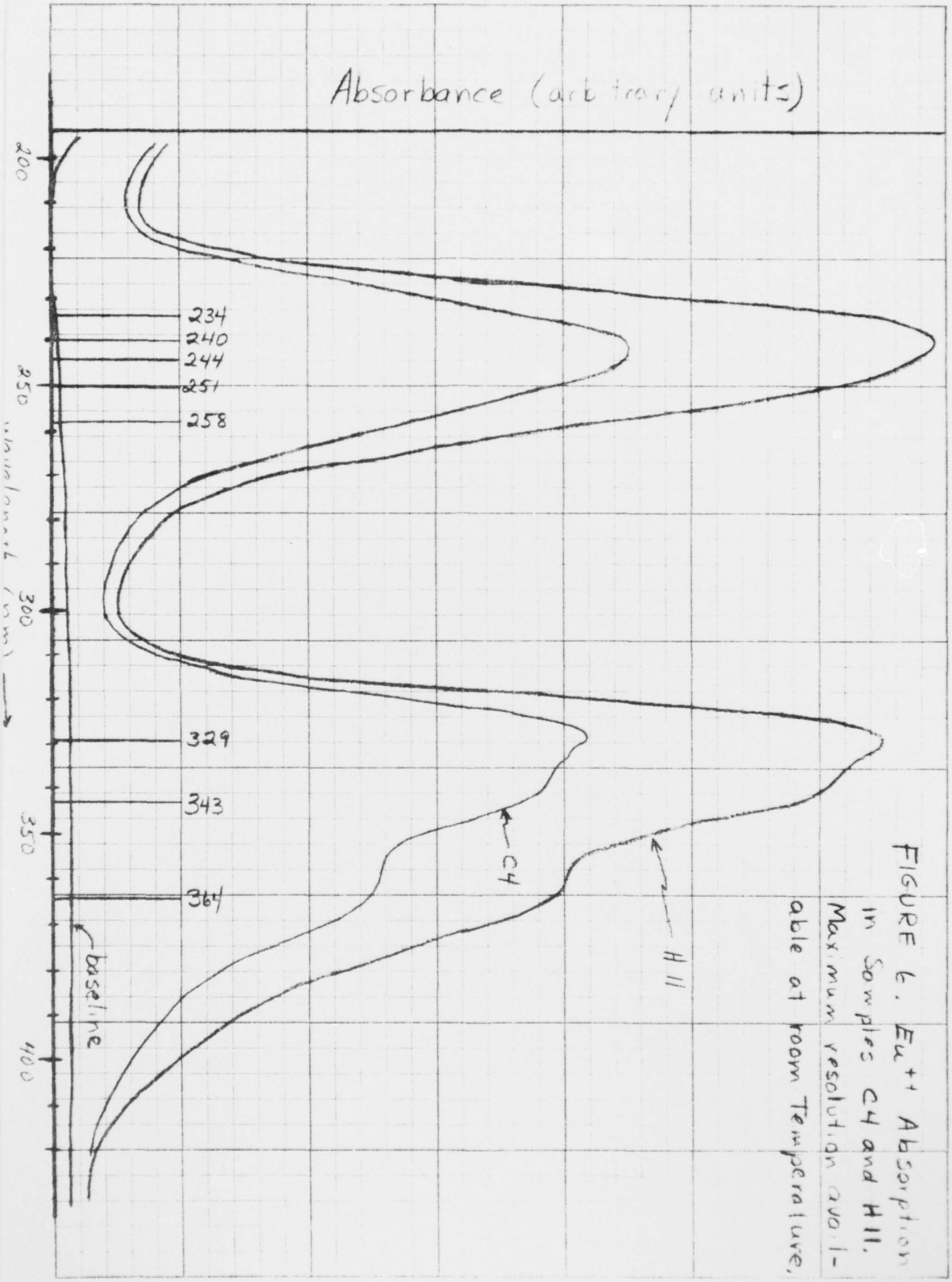


FIGURE 6. Eu++ Absorption in Samples C4 and H11. Maximum resolution available at room Temperature.

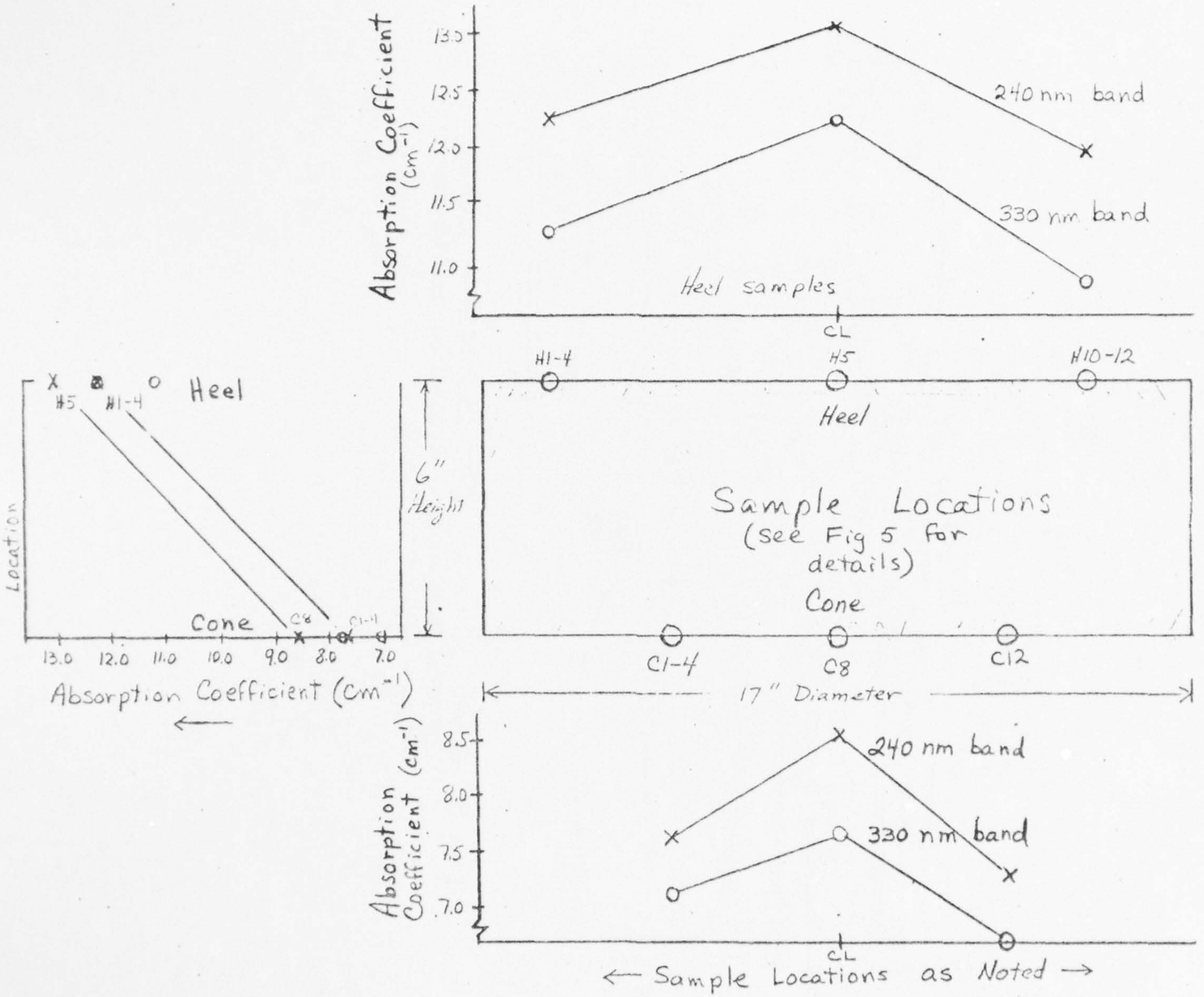


FIGURE 7. Eu^{++} Absorption Coefficient vs. Sample Position

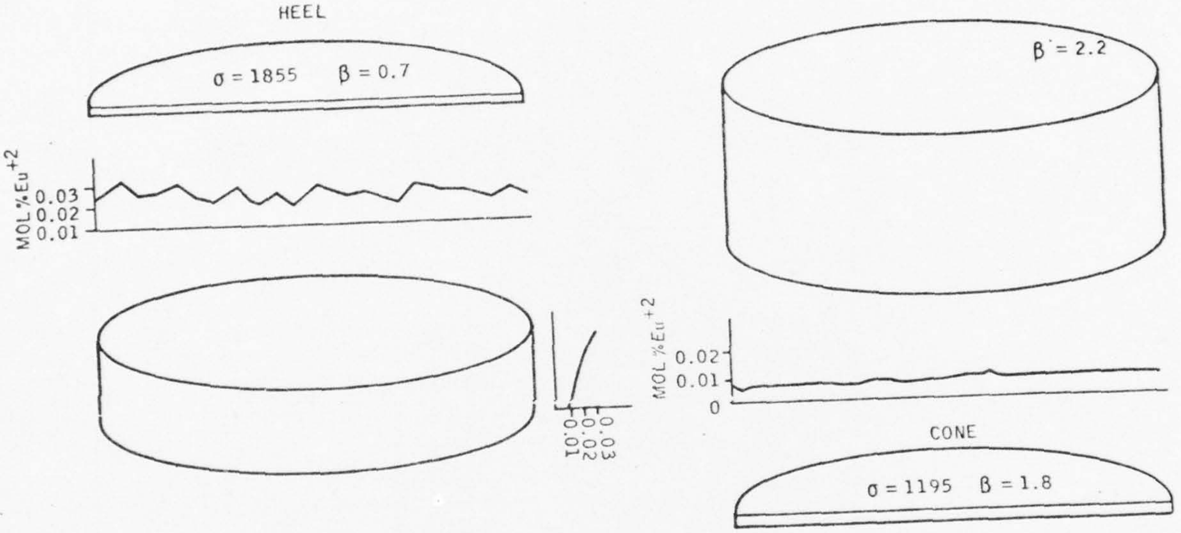


Figure 8 Schematic of Ingot No. 95 (KC. 01ECH95) Showing Chemical Composition, Absorption, and Mechanical Strength Data. From ref. 2.

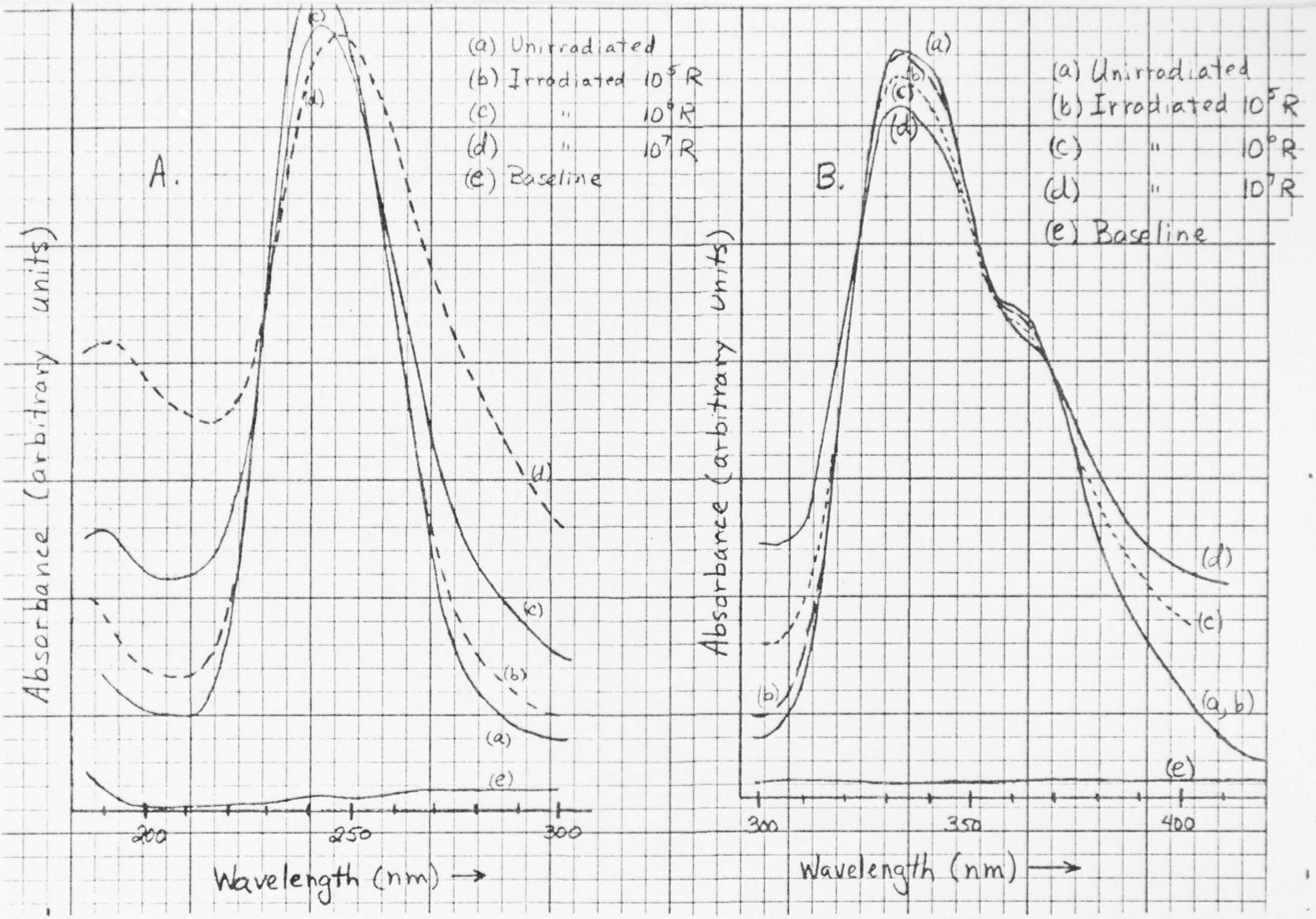


FIGURE 9. Eu^{++} absorptions in sample H10 as a function of radiation dose.

FIG 10

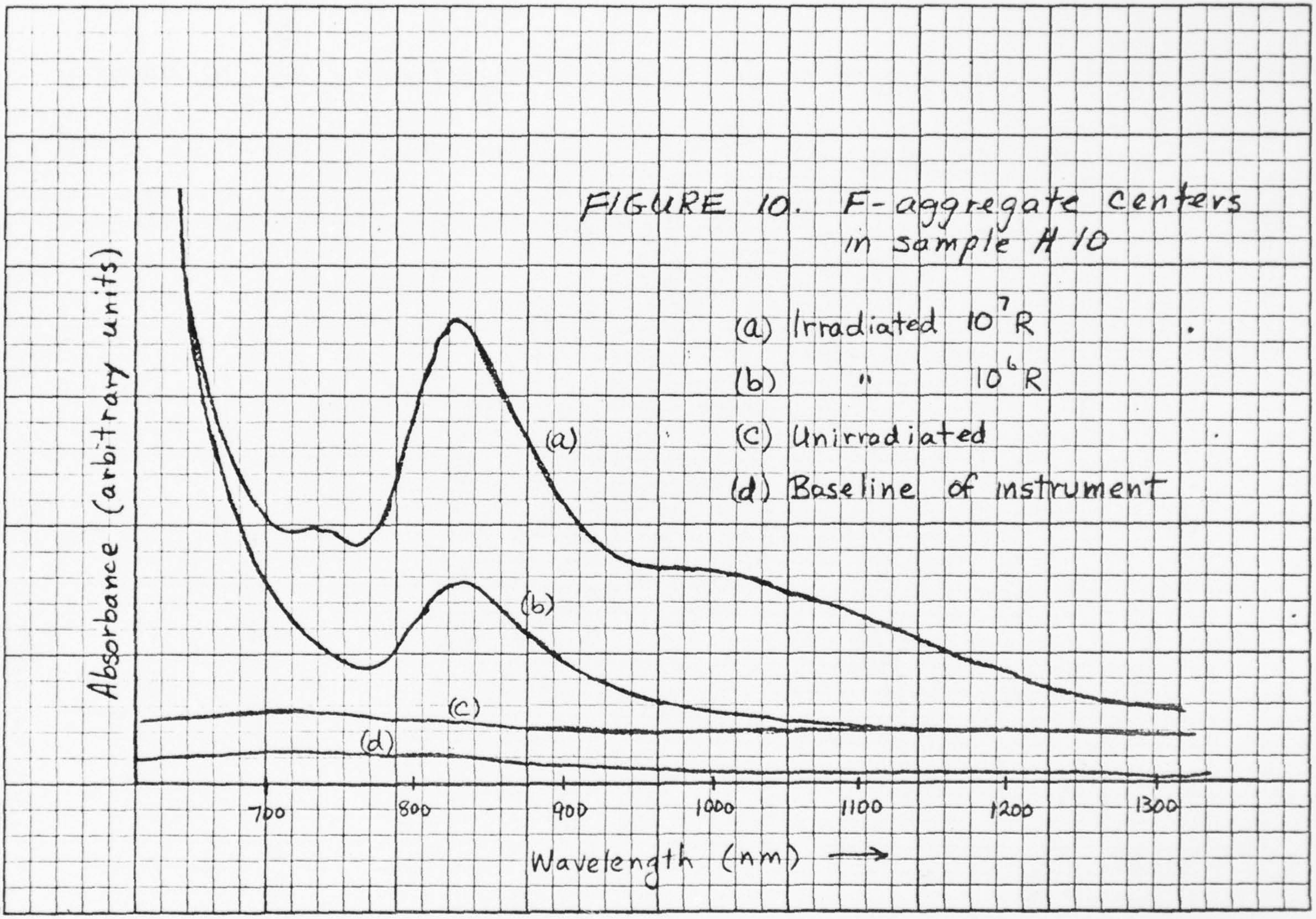


FIGURE 10. F-aggregate centers in sample # 10

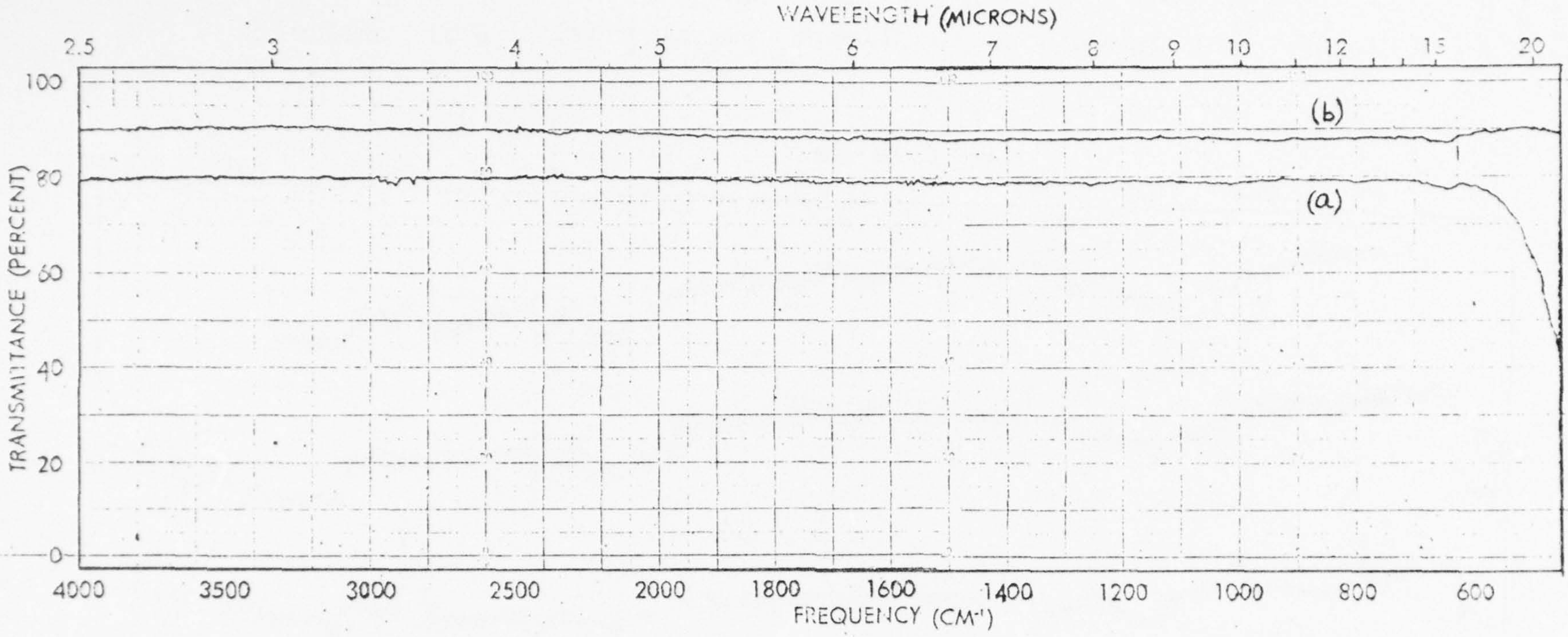


FIGURE 12 A. (a) Infrared absorption in sample H 10 after irradiation to 10^7 R.

(b) Instrument baseline.

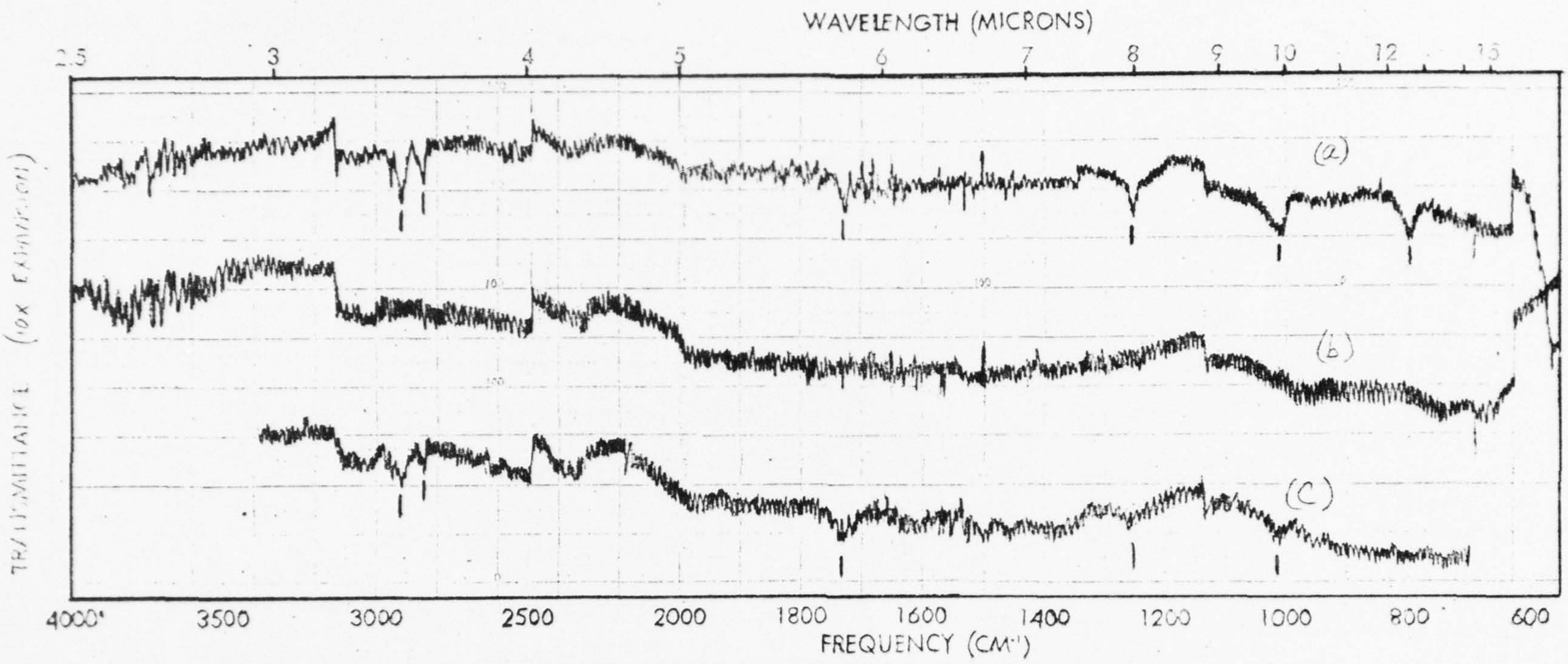


FIGURE 12B. (a) Infrared absorption in sample H 10 after irradiation to 10^7 R. 10X expansion.
(b) Instrument baseline, 10X expansion.
(c) Infrared absorption in sample H12, no irradiation. 10X expansion

Note

To reduce the length of this report, you may reduce figures & put 2 to a page; also tables I & II can go on one page.

I have made Xerox reductions to give models (attached)

Figs 1 & 2
Figs 3 & 4
Figs 12A & 12B
Tables I & II

Also, Fig 9 should be reduced to fit on a page (sample attached).

T O Stoebe

Table I. Experimentally observed absorption frequencies of several impurity ions in KCl crystals and the estimated impurity concentration to produce $\approx 10^{-4}cm^{-1}$ at the CO_2 laser frequency (from Ref. 27)

Ion	Frequencies(cm^{-1})	Conc.(ppm) for $\beta(943cm^{-1})=10^{-4}cm^{-1}$
H^-	500	> 10
OH^-	3640	> 100
SH^-	2590	> 100
CN^-	2089	100
BO_2^-	590,1972	10-100
CO_2^-	1696	10-100
N_3^-	642;2049	100
NCO^-	629,1232,2169	10
NO_2^-	805,1290,1329	0.1
NO_3^-	842,1062,1396	10
CO_3^{2-} a	680-720,883,886,1058,1064,1378,1416,1488,1518	10
BO_3^{3-}	736,949(w),1222	1-10
SeO_3^{3-}	broad738-850	1-10
HCO_3^-	589,672,713,840,971,1218,1346,1701,3339	0.01
BH_4^-	1143,2321	10
SO_4^{2-} a	630,978,1083,1149,1188	0.03
CrO_4^{2-} a	862,890,930,941	< 0.1
SeO_4^{2-} a	834,860,909,923	< 0.1
MnO_4^{2-} a	899,909,914,925	< 0.1
NH_4^{2+}	1405,3100	10

[a] Exact frequencies and intensities very dependent on presence of compensating M^{2+}.

Table II. Color Centers in Alkali Halides

Description of Defect	Notations used
Anion vacancy (uncharged)	F^+, α
Anion vacancy with one electron	F
Anion vacancy with two electrons	F^-, F'
Two adjacent F centers	F_2, M
Three adjacent F centers	F_3, R
Ionized M center	F_2^+, M^+
F center adjacent to monovalent cation impurity	F_A
F center adjacent to divalent cation impurity	F_z, Z
Intersitial halogen atom	
(exists as a molecular ion, two anions at one anion site)	H, Cl°
Self-trapped hole (one anion at one anion site)	V_k

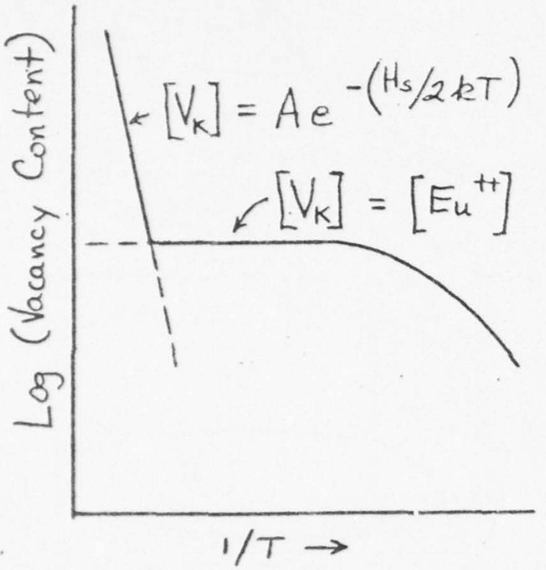


FIGURE 1. Temperature dependence of Vacancy concentration.

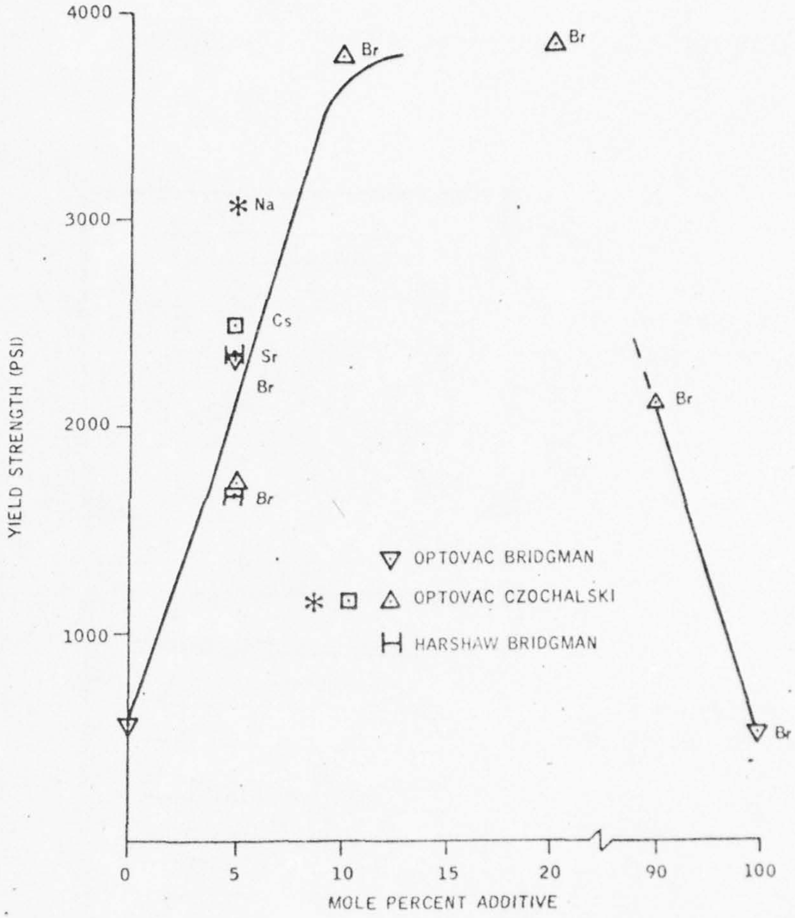


Figure 2. Influence of Na, Cs, Sr and Br Alloying Additions on KCl
(From Ref. 16)

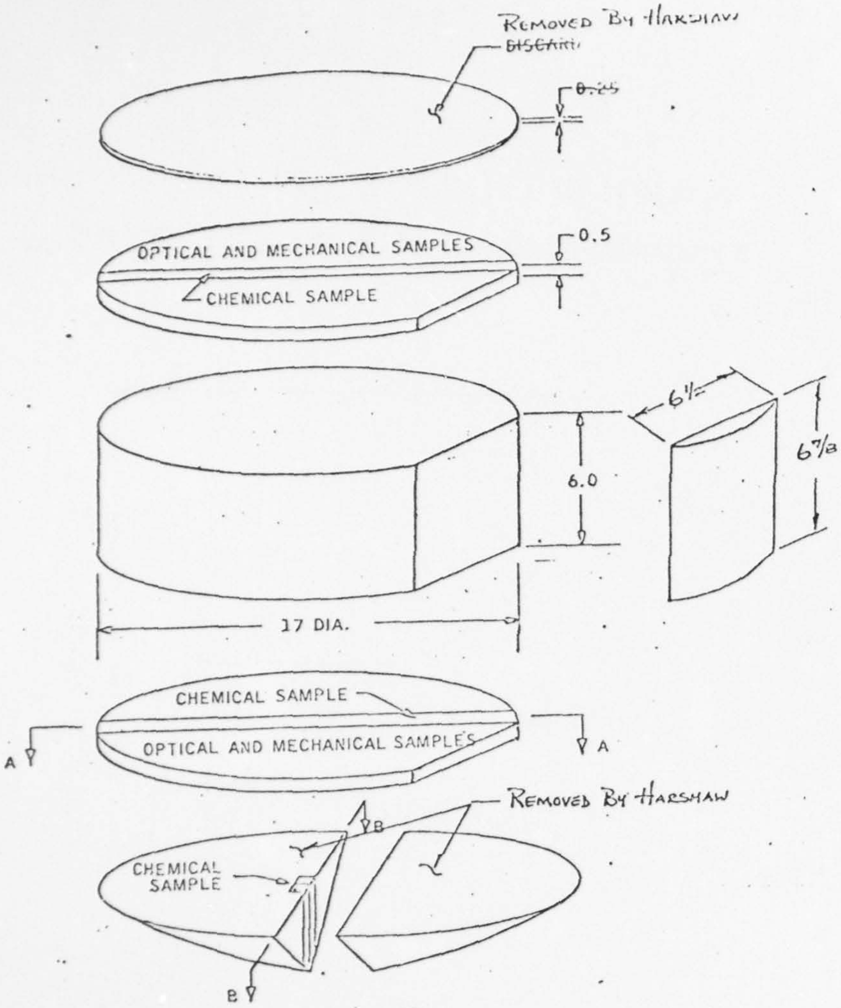


FIGURE 3. SECTIONING PLAN FOR KC.01ECH97

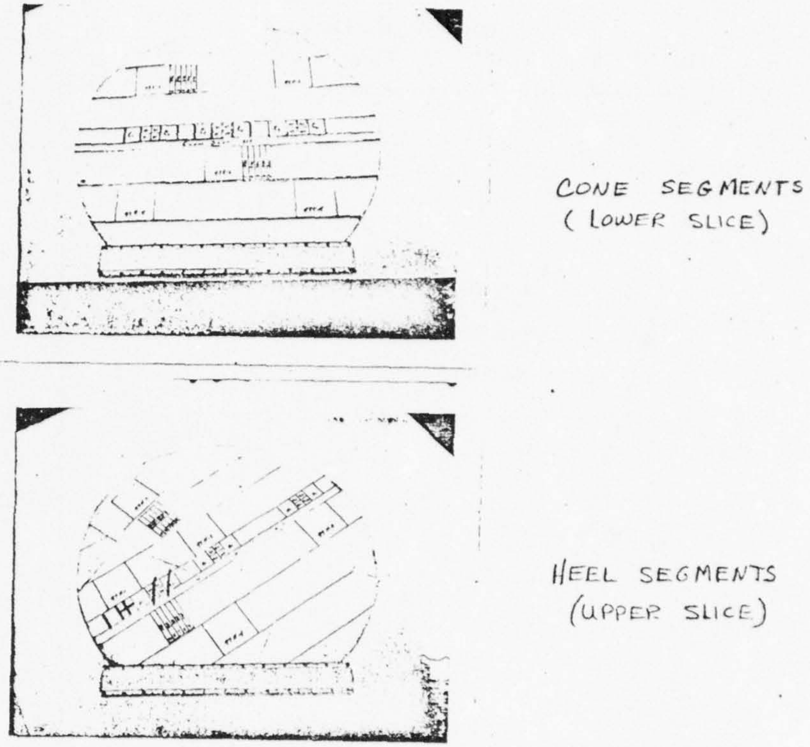


FIGURE 4: Sample Segments From KC.01 ECH 97

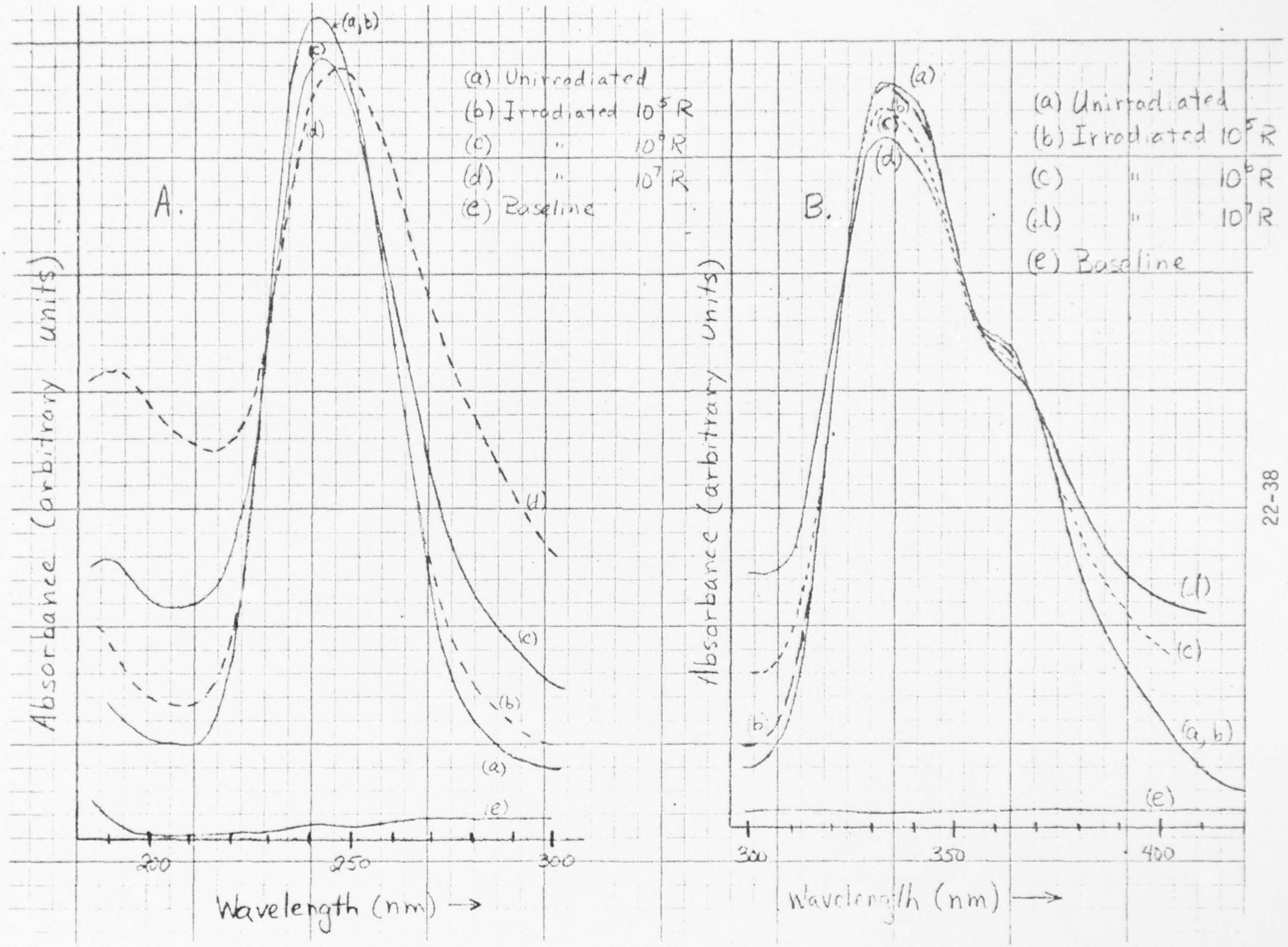


FIGURE 9. Eu^{++} absorptions in sample H10 as a function of radiation dose.

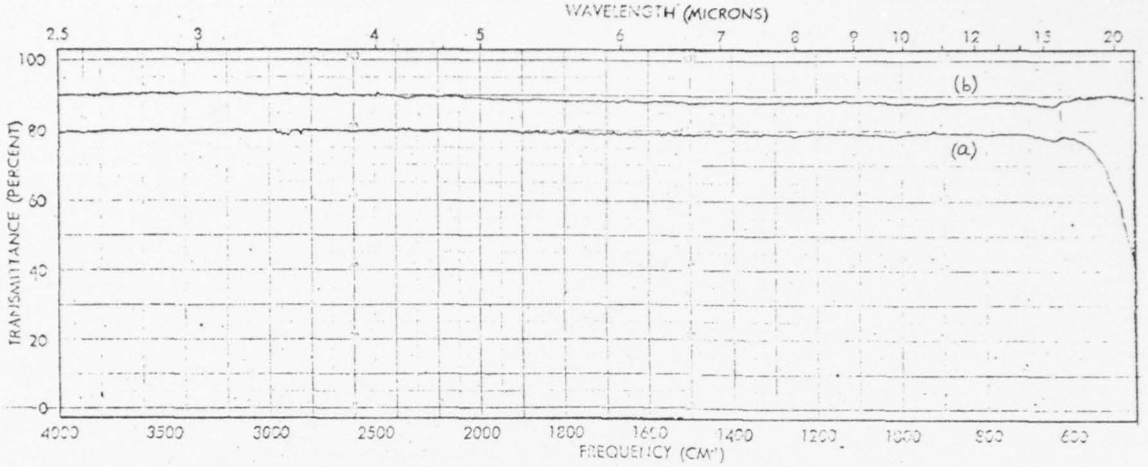


FIGURE 12 A. (a) Infrared absorption in sample H 10 after irradiation to 10^7 R.
(b) Instrument baseline.

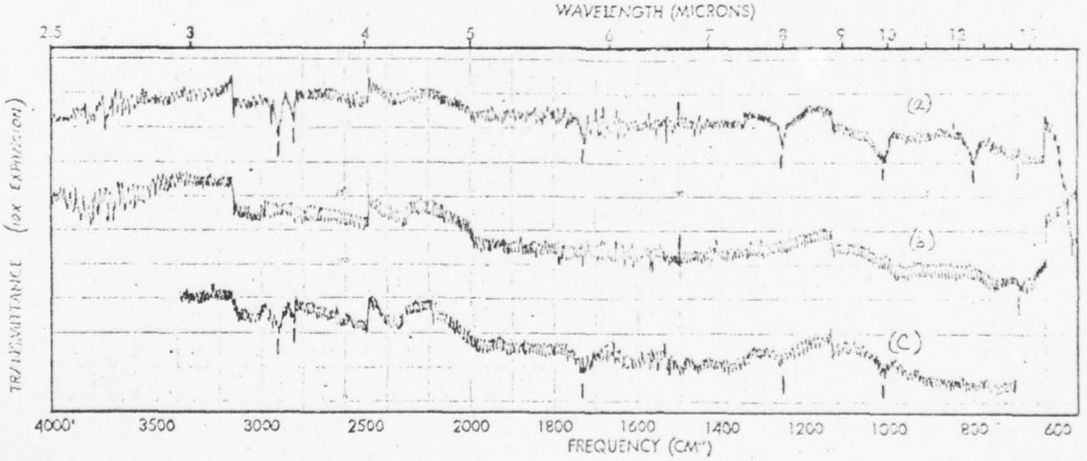
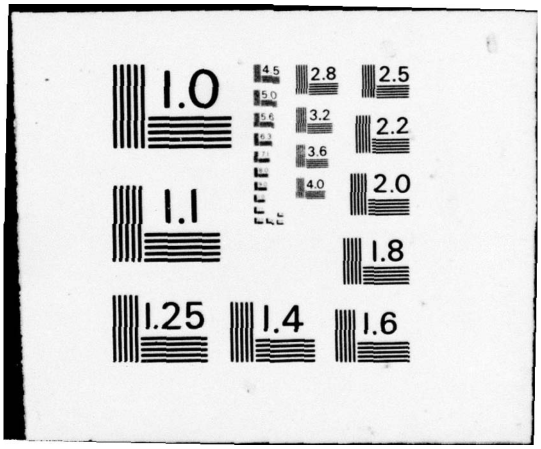


FIGURE 12 B. (a) Infrared absorption in sample H 10 after Irradiation to 10^7 R. 10x expansion.
(b) Instrument baseline, 10x expansion.
(c) Infrared absorption in sample H12, no irradiation. 10x expansion

AD-A031 017 AUBURN UNIV ALA SCHOOL OF ENGINEERING F/G 5/2
1975 USAF/ASEE SUMMER FACULTY RESEARCH PROGRAM.(U)
SEP 75 J F O'BRIEN F44620-75-C-0031
UNCLASSIFIED AFOSR-TR-76-1113 NL

7 OF 7
ADA031017

END
DATE
FILMED
11 -76


1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR - TR - 76 - 1113 ✓	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SUMMER FACULTY PROGRAM IN ENGINEERING AND SCIENTIFIC RESEARCH ✓		5. TYPE OF REPORT & PERIOD COVERED Final
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) J. Fred O'Brien, Jr		8. CONTRACT OR GRANT NUMBER(s) F44620-75-C-0031 NEW
9. PERFORMING ORGANIZATION NAME AND ADDRESS Auburn University Engineering Extension Service Auburn, Alabama 36830		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 9768-02
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (XOP) 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE Sep 1975
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release;
distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Summer Faculty Research Program

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report represents the first years effort of an annual ten week summer research program conducted by university faculty members at selected USAF Systems Command laboratories. Program objectives are: 1) Further the research objectives of the AFSC labs; 2) Enhance the research interests and capabilities of engineering educators; 3) Stimulate continuing relations among participating faculty members and their professional peers in the AFSC labs; 4) Form the basis for continuing research of interest to the Air Force at the participant's institution.